Alma Mater Studiorum – Università di Bologna

DOTTORATO DI RICERCA

INGEGNERIA AGRARIA

Ciclo XXI

Settore scientifico disciplinare di afferenza:
Costruzioni rurali e territorio agroforestale (SSD AGRO/10)

TITOLO TESI

THE ASSESSMENT OF CHANGES IN RURAL LANDSCAPE AND THE ANALYSIS OF THE DRIVING FORCES. PROPOSAL OF A SPATIAL MODEL FOR THE RURAL BUILT ENVIRONMENT

Presentata da: Minarelli Francesca

Coordinatore Dottorato | **Relatore**

Prof. A. Guarnieri | **Prof.ssa P. Tassinari**

Esame finale anno 2009

Keywords:

landscape change
rural built environment
generalized linear model
driving forces

Acknowledgements:

Thanks to my family and friends, for support. Thanks to Prof. P. Tassinari, Dr. D. Torreggiani, Dr. S. Benni and Prof. A. Guarnieri, of the Department of Agricultural Economics and Engineering, University of Bologna, and to Prof. N. Zimmermann, of the Swiss Federal Research Institute of Zurich, for the opportunities.

INDEX

Index

I INTRODUCTION & AIM OF THE STUDY

I. INTRODUCTION AND AIM OF THE STUDY

I.1. CONCEPTUAL BACKGROUND

Several definitions have been attempted in order to express the concept of landscape. The complexity is due to the polisemantic meaning of "Landscape" to which can be associated both ecological and aesthetical concepts. The word landscape, first recorded in 1598 derived by a painters' term borrowed from Dutch during 16th century when they were becoming master of landscape gender. In Dutch the original *landschap*, meant 'region, part of land', acquired the artistic sense which was turned into the English meaning of 'a picture depicting scenery on land'. Currently the most prominent official definition of landscape is provided by European Council and states that landscape intends an area, as perceived by people, whose character is the result of the action and interaction of natural and/or human factors (European Council, 2000, art. 1). The idea that landscapes evolve through time as a result of being acted by natural forces and human beings, that the landscape forms a whole complex, whose natural and cultural components are taken together not separately, appears in the definition. The understanding of the important role of the landscape on cultural, ecological, environmental and social fields and hence on the quality of life for people, increased social concerns regarding all issues mining the identity of the landscape. It appears clear that man's work is responsible in large amount of transformation dynamics which are the result of agriculture, forestry and industrial activity, regional planning policies and also at a more general level, changes in the global economy. Rural landscape has been defined by Sereni in 1962 as the form impressed by man and his production over the centuries consciously and systematically on the territory; it is the result of a transformation process implemented by men. Such concepts are also recall in the declaration issued at the rural development conference organised in Cork 1996. In this occasion the European Union stated that a quarter of the population lives in rural areas and account for more than 80% of the territory of the European Union. They are characterised by a unique cultural, economic and social fabric, an extraordinary patchwork of activities, and a great variety of landscapes (forests and farmland, unspoiled natural sites, villages and small towns, regional centres, small industries), and believes that rural areas and their inhabitants are considered as a real asset to the European Union hence recognised the important role of agriculture as interface between population and environment recalling the prominent function of farmers in the preservation of natural resources.

Throughout ages human impressed obvious traces of those interventions on the territory. Since Roman colonization to nowadays lands have been strongly modified in their land use for example from marshland to agricultural land, in their aspects creating roads network reticules and other landmarks like the *Roman centuriation* that represented the seeds for further settlement development (fig.1).

Figure 1 – Landscape structure defined by the mesh of Roman centuriation

However, while in the past, until the Second World War, changing dynamics of farming were likely to be gradual and discernible, after that time because of the conversion from a sustainable-oriented farming to a market-oriented farming, the development of residential settlements into agricultural space , and the progressive abandonment of low income rural area produced a rapid and abrupt transformation that is still leading to an ecological simplification and cultural erosion of the traditional rural landscape. In Flanders the concept of traditional landscape was introduced in 1985 by Antrop to carry out a study aimed at classifying geographical regions. Traditional landscapes have been defined as the landscapes which evolved during centuries until the fast and large scale modern changes started, with the introduction of technological power, thus corresponding to the Industrial Revolution time period. The modern impacts became really invasive after World War II with the economical boom that followed (see fig. 2). These changes deform the traditional structures, and thus their functioning, of the existing landscapes. In some places the traditional landscape was entirely substitute with completely new landscape. The modern landscapes are mainly characterised by uniform and rational solutions. Relicts of the traditional landscape structures still exist but in form of isolated patches which are more and more difficult to recognise.

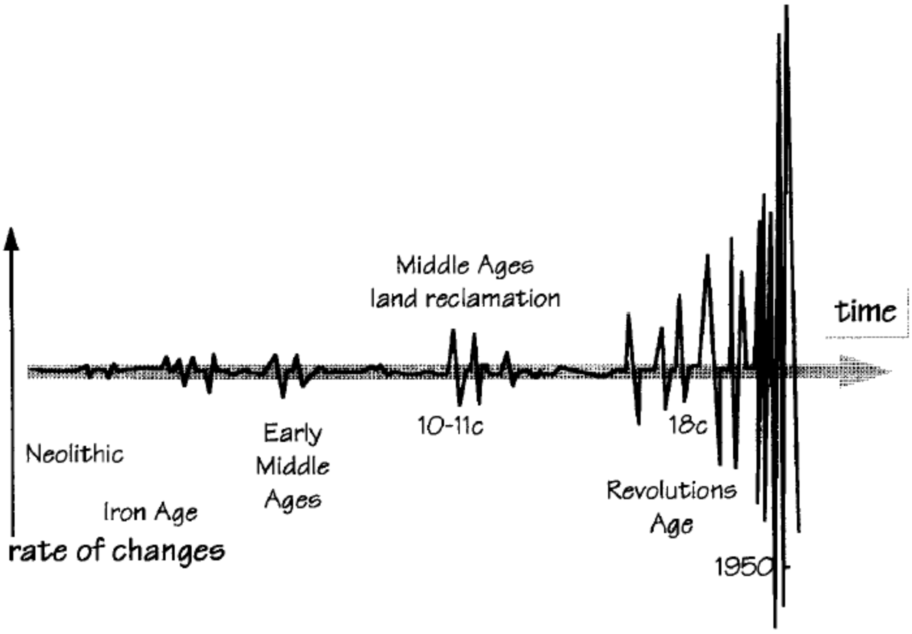


Figure 2 - Graph of magnitude and frequency of landscape evolution in Europe (Antrop 1997).

According to Antrop (1997) the traditional landscapes can be defined as those landscapes having a distinct and recognisable structure which reflects clear relations between the composing elements and having a significance for natural, cultural or aesthetical values and literal in their approach. Consequently such definition is referred to these landscapes with a long history, which evolved slowly and where it took centuries to form a characteristic structure reflecting a harmonious integration of abiotic, biotic and cultural elements.

I.1.2 The rural built environment

The rural built environment is the result of human action and represents one of the most important cultural element of the rural landscape. In its traditional forms it is an expression of humble culture, derived from agro-pastoral activity. The traditional artefacts are often made from local materials (wood, stone, earth, etc.) and have function of houses, stables, barns, local processing and storage of products, etc.. Technical solutions are essential, and practical to enable the farming activity and the use of all possible environmental resources. The advance of technology and modernity in building and agriculture have gradually changed the rural system by introducing campaigns works designed with the latest functional criteria, made using materials and equipment that are often inspired by models construction or industrial production. This has prompted to production development, but has also encouraged a gradual degradation of the landscape and the architectural quality of the overall agricultural areas, and the inclusion of facilities and activities within rural areas and that increase their impact on the environment. Emilia Romagna region, located in the north-eastern pert of Italy is on of the region mostly characterized by a prevailing agricultural production vocation and were transformation dynamic intensively occurred. It has been recognised as including many interesting cases of aggregation and urban sprawl by numerous authors, including Ingersoll (2001, 2002) and Indovina et. al. (2005) and therefore assumed as area of investigation of such phenomena. The traditional farm in Emilia Romagna was organized in residential building and different rural annexes functional to agricultural activity such as stables and haylofts. The whole arranged complex of buildings was named *rural court* (fig.3).

Figure 3 - Rural court

In these last decades rapid transformation due not only to the introduction of new technologies in the production process but also to new agricultural polices and new economic dimension generated deep modifications in the agricultural sector and in particular in the role that agricultural activity has assumed in the production system. The abrupt conversion happened after second world war shifting from a subsistence agriculture to an industrial agriculture characterized by a high use in pesticide and mechanization and so by a rearrangement of farm lands to allow such interventions provoked a general simplification of the rural landscape. Elements of rural landscape losing their original functionality related to rural context were quickly modified or substituted with more suitable uses for modern agricultural practices or new land use destinations. Rural built environment as component of rural landscape has been strongly affected by such dynamics. Traditional farm buildings, due to their recognised unique elements, are important within the countryside. They may have visual qualities as part of the landscape, and they may have archaeological or historical value giving an indication of previous farming practices. In many cases buildings have been restored without considering such valuable elements and contrasting the original arrangement of rural landscape. Traditional farm buildings may no longer be required for their original use. Uses other than for agricultural purposes tend to be favoured, such as residential and business uses because they provide more benefits. The intense allocation of production factors related to urban activity in the rural spaces with a high level of antropic activity deeply modified original landscape. As a consequence there is an infiltration of urban patterns within agricultural spaces developing new assets where rural and urban patterns are both existent. Such interactions provoke several negative effects such as sprawl effect of settlement and a gradual conversion of rural spaces into urban. The subtraction of agricultural lands for urban purposes is becoming more and more intense attracting attentions of several academic fields at national and international level. Since landscape is an multidisciplinary topic because of the manifold resources involved, different disciplinary groups are taking in charge of the analysis of dynamics with the common purpose of preserving and protecting the landscape heritage. On this scenario planning policies represent a key tool by means of actuating a sustainable development of landscape. It is well known that the study of landscape changes is an essential stage for the promotion of conscious decision making in land planning. Institutions in charge of planning and programming should be aware of dynamics occurred in the past in order to prevent future scenario. On this purpose the comprehension of reasons responsible for landscape transformation results to be important elements to gain a better knowledge of possible present and future rural landscape dynamics. The study of changes in the landscape is a topic widely discussed in scientific community at national and international level. The analysis of changes occurred on various components of rural landscape with the main purpose of identifying reasons related to such transformation and to forecast future scenarios of evolution, represents an important tool and indeed a prerequisite as part of the landscape planning and programming. In relation to both these objectives techniques are frequently used, for modelling space, covering different time interval and adopting different methods depending on the specific field of analysis and resources of the landscape investigated. In particular the rural built system represents an important component of the landscape mosaic, whose type of development has a significant impact on the overall rural landscape evolution.

I.2 AIM OF THE STUDY

The specific aim of the presented study is to propose a methodology for the development of a spatial model that allows the identification of driving forces which mostly have influence on building allocation.

It is well known that the study of landscape changes is essential for conscious decision making in land planning, and analytical methodologies aimed to such study are part of a topic broadly developed by the scientific community. The review of the scientific literature carried out pointed out the importance of focusing on criteria for analysing the trends in the rural built environment and on landscape modelling techniques which can be developed for such purposes, but also a general lack of modelling methodology for the study of rural built environment changes.

There are mainly two types of transformation dynamics that have been recognised as affecting the rural built environment: the conversion of rural buildings and the increasing of building numbers. The present study decided to focus on the analysis of the latter dynamic for reason related to the concerning issues brought about by the expansion of built up areas in rural landscapes. Moreover another distinction must be done regarding the dynamic affecting the increasing of building number. Similarly to what occurs in land use change analysis, the investigations on rural building expansion can be carried out following two different approaches: the identification of the rate of expansion of rural buildings or the identification of their spatial allocation. Each of this option generates a very different outcome. In the first case the answer addresses to the question which rates the expansion are likely to progress, i.e. the quantity of expansion, and in the second case where the expansion is likely to take place. The allocation analysis is based on spatial analysis of the complex interaction between rural built system, socio economic condition and biophysical constraint locations of land use change and requires the identification of the natural and cultural landscape attributes which are considered the spatial determinants of change. Conversely, the rate or quantity of change are driven by demand for land-based commodities (Stephenne and Lambin, 2001) and in the case of land use change models they, are often modelled using economic framework (Fisher and Sun, 2001). This means that driving forces which control the rate of changes operate at higher hierarchical levels hence that they often involve macro-economic transformation and policy changes (Lambin et al., 1999).

Understanding the spatial distribution of data generated from events that occur in space constitutes today a great challenge in many research fields because it results having direct implications in planning activities. In fact the importance of recognising driving forces that operate in the selection of land suitability for building construction represents a much more useful tool than the detection of rates. In the decision making process, actuated for planning purposes, it is fundamental to know driving forces responsible for the allocation of new built up-area in order to contrast irrational expansion of building across landscape and to prevent urban sprawl and landscape fragmentation.

Hence the design of the model involved the identification of predictive variables (related to geomorphologic, socio-economic, structural and infrastructural systems of landscape) capable of representing the *driving forces* responsible for landscape changes. The response variable was represented by a binomial function indicating presence or absence of buildings in a general location of the study area.

II LITERATURE REVIEW

II. LITERATURE RIEVIEW

Scientific literature provides a wide range of studies that deal with landscape changes and provides different socio-economic or ecological approaches.
According to Antrop (2000), landscape should be considered as holistic, relativistic and dynamic. Landscape is dynamic in the sense that the nature of the composing elements changes under diverse impulse including human action. In particular, over the last century the evolving of landscape became increasingly affected by human behaviour determining abrupt changes such as in land use, species distributions, land morphologies. For this reasons, efforts from scientific communities are mostly tuned on transformation analysis. In literature there are many application examples such as Turner (1989), Li et al.(1993) and Dunning et al. (1992) that used various landscape indexes to measure landscape structure under different conditions to assess entity of change. They reveal that such studies are not sufficient to explain some research questions, there is also a need for models.
Models of landscape changes have been reviewed by Shugart and West 1980; Louks et al.1981; Weinstein and Shugart 1983; Shugart 1984; Risser et al.1984; Shugart and Seafle 1985 and Baker 1989. A variety of criteria could be used to distinguish models of landscape change. In 1987 Baker proposed an approach for the study of landscape change based on these following two criteria: (1) the level of aggregation, and (2) the use of continuous or discrete mathematics. The level of aggregation criterion refers to the level of detail with which the landscape change process is modelled. Following this approach, Baker classified landscape models into three categories: (1) whole landscape models, (2) distributional landscape models, and (3) spatial landscape models. Whole landscape models focus on the value of a variable or several variables in a particular land area. Distributional landscape models emphasize changes in the distribution of lands cover types. Spatial models, on the other hand, use the location and configuration of landscape elements in projecting change and explicitly produce maps of these changing spatial configurations. Whole landscape models describe the characteristics of a certain landscape; distributional models examine or predict changes of landscape distribution over time; and spatial models determine where the changes might occur. Markov model are probably the most commonly used distributional models. Spatial landscape models take into account the spatial configuration of the landscape phenomena and recently this approach thanks to advancement with GIS technologies found wider application in landscape modelling. Spatial model are classified in mosaic landscape model in which change in mosaic of individual subarea is modelled and element landscape

model in which change in individual landscape elements is modelled. The main difference is that in the first case the change is measured in subarea by which landscape is divided. The phenomena for example the amount of houses are taken within a equal portion in which landscape is subdivided like a mosaic shape. In this way it's known the amount of change per subarea. The element model landscape focus on the response of individual element to the spatial configuration. Such type of model find application for the study of organism in habitat distribution model and they use grid cell or vector-based mapped organism location. Literature shows some cases in which this approach is applied for modelling individual landscape element since analogies can be existent between these two entities. Especially where biotic interaction strongly control the character of landscape element or create patches, individual organism model may be more appropriated then mosaic model.

Rural landscapes often absolve complex and competitive demands of society. For example, they are used by people to generate income (ex. agriculture, mining, and tourism), to provide a living space, and to provide quality of life (clean water, recreation, and social activities). Rural landscape has created indeed a new ecosystem and a better understanding of the ecology of the landscape and the whole complex of components is needed to provide sustainable future rural landscapes. Nichol G.E. et al. (2005)proposed a classification model braked into model domains and subdomains: socio economic, biophysical (land-use, ecology, soil and water hydrology, hydrogeology, agricultural production, environmental pollution and nutrient flow), and generic and integrative (planning support system, visualization, decision support system, risk assessment, climate change). The subdomain of land use and land cover change models have become increasingly common because of the dramatic expansion of developed area that provokes fragmentation of landscape and urban sprawl. Early models of land-cover and land-use change were essentially non-spatial (e.g., Johnson, 1977). Later models incorporated a set of explanatory variables that might themselves be spatially patterned. The probability of development at some pixel i is a function of a vector x of explanatory variables.

More complicated functions in x can contain information on state of nearby pixels, given the model attribute of cellular automata (Neumann, 1966; Wolfram, 1984; Hogeweg, 1988). The Spatially explicit land use/cover change models can be divided in three categories: simulation, estimation and a hybrid approach that includes estimated parameters with simulation. Many of the simulation models are based on a cellular automata approach. These are a class of mathematical models in which the behaviour of a system is generated by a set of deterministic or probabilistic rules that set the state of a cell based on the states of neighbouring cells.

Individual cell states are updates on the base of neighbour cell states in the previous time period. Such models have been used widely to study process of urban growth (Wu and Webster,1998; Clarke et al.,1997; White et al.,1997; White and Engelen, 1993; Batty et al., 1989). The progress made in technology using GIS technology and remote sensing have made possible the incorporation of the space-time components at different levels of resolution within the model enhancing the modeling space for the study of the territory (Sklar & Constanza 1991). As the landscape is a complex entity and the use of dynamic modeling space was an important support in decision-making in the government of the territory for those involved in planning activities (Behan 1990), other are the benefits of its implementation, in particular the possibility of conducting simulations that can be displayed on the territory of study, to relate the phenomenon being studied with the surrounding factors and their distribution in time and space (Dunnning et al.1995). A wide use of this tools is common among geographer for developing spatially explicit empirical models. Examples from literature include Mertens and Lambin (1997), Andersen (1996), LaGro and DeGloria(1992). These focus on the aspect that deforestation whose data are acquired by means of remote sensing and represent the dependent variable are related to some explanatory variables also derived from the remotely sensed data.

In the majority of case studies, it is usual to fit a single model to the data describing land use change, then test that model and use it for projection into the future. additionally the focus in many models is the hypothesis testing and similar situation exists in ecology in relation to models of the distribution of plants and animal species, where for example a single model is developed from empirical data and use to predict distribution (Guisan and Zimmermann, 2000).

A wide range of studies is available for the modelling of urban area expansions. Referring to these an interesting approach is developed by W.F.Fagan et al., 2000. The stated hypothesis is that urbanization expansion must be considered as an ecological colonization process in which individual colonists (houses) occupy available space and influence subsequent development. From this perspective, processes of built-up area expansion are similar in many ways to the growth of plant population or the growth of animal species. Reasons that prompt to formulate such theory is mainly because urban growth have captured interested from geographers, environmental planners, and social scientist (e.g., Chapin and Weiss 1968), also developing urban growth models mostly by using neural network linked with GIS platforms, (Weisner and Cowen 1997) and heuristic optimization techniques (Densham 1991; Batty and

Densham 1996) but always lacking of an ecological approach. Hence, it is generally perceived the importance of incorporating also ecological principles into the analysis of human-dominated systems.

Another topic related to agricultural sector where literature review provides wide of contributions is agricultural land-use modelling. Agricultural land-use modelling has a long history, dating back to the agricultural location theory developed in the late 1700s. The last 10 years has seen an increasing attentions from many scientific communities perhaps because of the availability of cheaper and more powerful computer technologies, and the coming of natural resource management issues in many different fields which are often associated to the strong impact that agricultural activity has on natural environment. The need for useful models is increasingly apparent. A wide variety of methods has been developed that seek to have an impact on problem solving. Examples include spatial models of herbivore (Coughenour, 1991), crop distribution models (Carter and Jones, 1993), agricultural location models to explain deforestation patterns (Chomitz and Gray, 1996), explanatory Markov rule-based models of land-use dynamics in a watershed (Stoorvogel, 1995), and a whole array of statistical and simulation models that contain
some spatial components to study land-use and deforestation processes, (reviewed by Lambin, 1994).

Although a wide literature review that faces issues related to rural environment, changes occurred on rural built environment have never been treated as specific objective of study, except for one case reported by R.Aspinall, 2004. This illustrated a case study carried out in Montana, where rural housing change locations are uses as parameter of land use change. The design model is used to develop a series of different models reflecting drivers of change at different periods in the history of the study area.

It can be said that there is a general lack in scientific literature of models for the study of rural built environment changes, that instead represent an important component of rural landscape recently affected by strong transformation dynamics.

III MATERIALS & METHODS

III. MATERIALS AND METHODS

III.1 STUDY AREA

To test the model developed by following the defined methodology, it is necessary to apply the model to a case study. Hence to select a study area as a target, together with a time range that are representative of dynamics under study where to carry out the calibration process. The significance of the landscape for the phenomena under study motivates the adoption of a study area located in the northern part of Italy within Emilia Romagna Region named New District of Imola (a description of the study area is reported in paragraph III.1). The time range selected for testing the model is the interval between 1975 and 2005 since such period of time is extremely meaningful for transformations occurred on rural built environment.

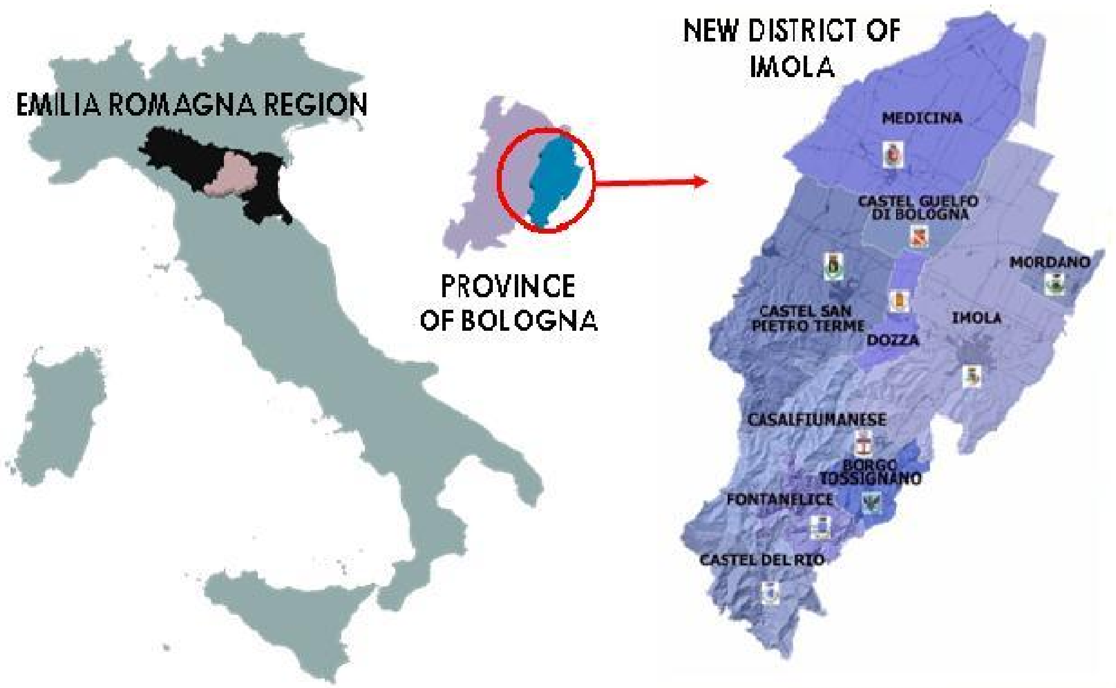


Figure 4 - New District of Imola.

III.1. 1 NEW DISTRICT OF IMOLA

The study area is located in the eastern part of Bologna Province within the Emilia Romagna region. The extension is approximately 787 Km^2 with a population of 125.000 habitants and it is comprehensive of ten municipalities which are characterised by sharing similar geomorphologic conditions and historical past events.

The landscape is heterogenic: flat lands extending to the northern part gradually turn into hills and relieves moving toward the southern part. For each of these landscape scenarios small or medium municipalities are sprawl across the territory.
Existent municipalities find their origins in pre-existent ancient settlements and they still maintained some residuals of traditional rural activity that were developed in accordance with the type of environment. Along this heterogeneous landscape, different typologies of rural settlements are well represented, as expression of rural architecture adaption to the agricultural activity at different environments. Significant geomorphologic and landscape diversification, make possible to observe structured dynamics for all the main land use systems and have an overall depiction of rural built environment and keep trace of transformation occurred in different types of landscape.

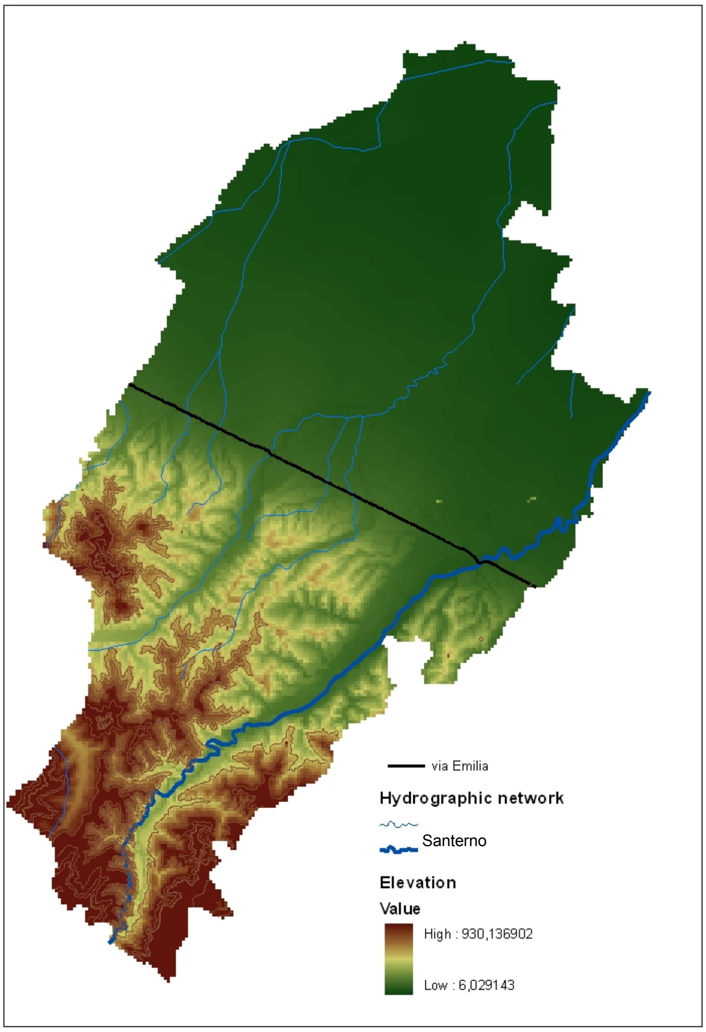


Figure 5 - Study area morphology

The elevation digital model allows to distinguish in the study area three main thresholds at 50, 300, 600 and 900 m above sea level, which respectively individuate three regions: plain area

(46% of the entire area), hill-foot area(38% of the entire area), hilly area (15% of the entire area) and low mountain (1% of the entire area) (fig.5). The most important river that crosses the area is represented by the Santerno whose valley follows a flow direction from the southwest to northwest between Apennines and foot-hill area. The landscape of the southern part of the study area is portrayed by hills that sometimes is characterized by marginal agriculture and uncultivated areas, forests.

According to a research study reported by Tassinari et al (2007), the province of Bologna, between 1976 and 2003, showed an expansion of the inhabitant urban centres and an intensification of the urban fringe starting from Bologna centre and extending along the main roads. Such expansions are particularly stressed along the most important road which is the via Emilia (fig. 7).

Others important sprawls occurred also across flat lands followings the surrounding of Santerno and Reno rivers. Also a fusion between Bologna's urban territory and the one of the municipalities in the immediate suburbs, with the consequent formation of a compact urban area has been observed, and in particular its ultimate boundary is larger than the municipal boundaries of the main city and embraces farm fields and fringe area. The growth of towns increase along via Emilia until the city of Imola.

In the flat part that still follows the structure defined by the mesh of Roman centuriation it can be easily recognised the via Emilia, which ideally divides the study area into a northern flat and a southern hill part. The study area has been object of a classification resulting from the analysis of elevation joint to agro forestry land use suitability defined by Klingebiel and Montogomery, 1961.

III.1.2. HISTORICAL BACKGROUND

The first significant actions on the territory under study, occurred in the Roman era in particular with *centurition* whose form impressed the landscape and maintained over time representing a base for various subsequent transformations. Also Marinelli in the XIX century states that, over 100,000 hectares of plains of Italy, apparently among the most fertile, reveal the obvious division of the ancient land. Roads, canals and ditches follows a pre-territorial division, and a network of roads and drains that can not easily be changed (fig.6)

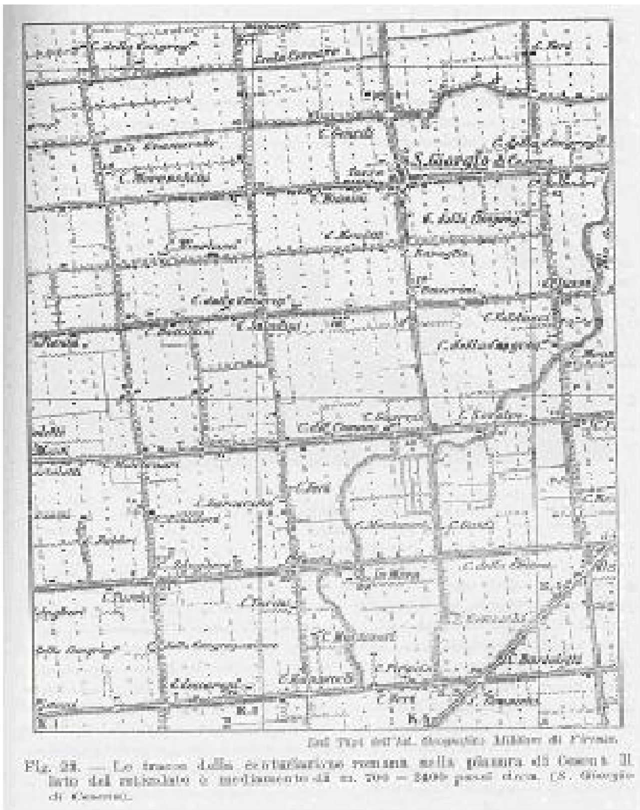


Figure 6 - Landscape structure defined by the mesh of Roman centuriation.
The side of the mesh was set to 700 meters.

In Roman ages the capacity of modifying the landscape was very intense, resulting a strong man-made landscape. The ancient road, *via Emilia*, built in the 187 B.C by Marco Emilio Lepido as a connection between Rimini, on the Adriatic coast, to Piacenza, on the river Po, passing through Bologna, Imola, Faenza has always represented one of the most important route of the Pianura Padana region. This vast country, by far the largest fertile plain in the mountainous peninsula, contained potentially its best agricultural lands, and offered to Romans the opportunity to expand enormously their population and economic resources by means of massive colonisation.

Figure 7 - Via Emilia in red.

During the second century B.C. various seed dwellings grew spontaneously in strategic geographic positions such as in proximity of Lamone and Santerno rivers that became: Forum Corneli (Imola) and Favente (Faenza). During medieval times, after the fall of the Roman Empire, Romagna became part of the Byzantines domain, with the generation of a peculiar settlement systems inherited from Roman named as *fundus* and *massa.*

In that period, despite the loss of some centres, cities in Romagna continued to serve as a core for aggregations, increasing settlements. This trend is in contrast to what happened in urban centres of Lombard region, where number of settlements dropped down. An interesting dynamic in this period was the development of large properties (particularly ecclesiastical) with a consequent consolidation of funds.

Towns were built slightly before XI century and they inherited a basic structure that would have been never altered by subsequent interventions. After XI century in territories surrounding Imola the dynamics of population were attracted by the construction of churches and rural settlements scattered around the countryside. Between the middle of XIV century and the beginning of XV Imola was affected by a strong expansion of the urban perimeter because of a wide immigration that placed the city among the twenty most populous of the peninsula. Until XIV there were not significant events that introduced changes in Romagna landscapes.

Among the XIV-XV century the introduction of the asymmetrical plow, determined a change in the form of fields, no longer square but rectangular lengthened.

During XIV-XV century the landscape is strongly characterized by a form of promiscuous crop called *Piantata* (*arbustum gallicum*) (fig.8), formerly known in Etruscan era, with the main function of drainage. These lines defined the shape and size of farms, paths and fields. The transformation of *Piantata* began during XVIII century with the introduction of new crops such as corn and hemp crops requiring highly exhausting new agricultural practices with great movement and transportation of soil and resettlement of hydraulic structures. Not only lines of *Piantata* and drains contributed in the shape of rural landscape but also trails that were built for transportation and communication uses.

Figure 8 - Piantata

The construction of a rail that privileged Bologna brought some crisis in the productivity of the Romagna plain. As rebalancing intervention some connections were created in Romagna such as a steam tramway Bologna - Imola, which integrated the Imola area in the space of Bologna.

After the building of numerous roads made from 1859 onwards, agriculture improved its production by about 1 / 3 of grain and rice. In this system increased the viability of minor roads having a key role in the overcome of farm fragmentation. Consequently countries were continually intersected by these roads forming a reticule representative of the huge land fragmentation level of these lands. The crisis of agricultural production because of the market of Asian and American products dropped the production of wheat which represented the main product of farms.

Beet crops at the beginning of 1900 was already successfully progressing in the rotation as agricultural crop roots and renewing. The advent of beet crop cultivation affects strongly the agricultural landscape. In these lands, between the national union and the first world war, landscape assumed profiles of long chimneys with sugar beet factories almost everywhere along the railway lines and spreading through cultivated fields.

Starting from 1920 the agricultural landscape of the land reclamation changed, following the introduction of wide orchard systems in Romagna, which for many years it was always maintained in extensive and not promiscuous intensive, with consequent relocation of facilities serving such activities as cold stores etc. According to the description of a typical rural landscape in Romagna at the beginning of 1900, reported by Oliva, the intensive form of cultivation determined a rectangular geometric shape of farm fields with a range of length between 500 msq to 8000 msq. This regular shape of fields evidences the remarkable mark of mesh created during roman centuries Such traces can be easily recognized along via Emila, on the border between Bologna and Modena, and between Lugo and Imola.

The second world war ended with the destruction of cities including infrastructure and hydraulic defence systems bringing a period of general crisis and poverty.

Several actions helped to improve the situation in this period of reconstruction. The main event was the opening of the domestic fruit market. This spurred the building of utilities related to fruit and vegetable market in Romagna.

The strong productivity growth not only affected the fruit sector but also brought radical changes at social and demographic levels resulting on abrupt modification of rural landscapes structure. In particular the growing use of mechanization in agriculture caused the abandonment of traditional cultural forms no longer adequate since an obstacle to new technologies, such as the destruction of the green lattice of *piantata*. Since the 60s, there has been a drastic reduction in agricultural area according to the abandonment of less productive lands and the increasing use of lands for urbanization and industrialization. Starting from a considerable fragmentation of extended land possession, those generated a rapid spread around small towns addressed preferably for building purposes.

III.2 *DEVELOPMENT OF THE MODELING METHODOLOGY*

The analysis of driving forces of landscape transformation dynamics represents one of the main research topic in many fields of study with the common main purpose of developing a sustainable planning activity aimed at preserving and protecting landscape resources.

Several ecological indicators have been developed in order to reflect a variety of aspects related to the ecosystem, including biological, chemical and physical, but they are not sufficient in order to understand reasons involved in changing process and to support simulation of future scenario. For such purposes more suitable analysis tools, such as statistical modeling techniques, that allows to understand driving forces and future scenario, needed to be introduced.

According to M.P. Austin (2002), three components are needed for statistical modeling: a model concerning the theory to be used or tested, a data model concerning the collection and measurement of data, and a statistical model concerning the statistical theory and methods used. In the following paragraphs are processed these steps that lead to the definition of the most suitable model for the comprehension of reasons related to built environment allocation within rural and periurban area. The model developed by following the methodology is applied to a case study to test the validity of the methodology. In particular the model has the specific aim of testing the existence of a cause-effective relationship between some possible factors affecting expansion dynamics and the increase of the built environment in term of their allocation. Hence the first step in the development of the model consists in identifying driving forces responsible for rural built environment expansion. This assumption also called hypothesis is formulated stating the existence of a cause effective relationship between selected driving forces and building allocation process.

III.2.1 THEORETIC MODEL

A different combination of factors in various parts of the territory generated favourable or less favourable conditions for the building allocation and the existence of buildings represents the evidence of such optimum. Conversely, the absence of buildings reflects a combination of agents which is not suitable for building allocation. Presence or absence of buildings can be adopted as indicator of such driving conditions, since they represent the expression of the action of driving factors in the land suitability sorting process. The existence of correlation between site selection and hypothetical driving forces, evaluated by means of modeling techniques, provides an evidence of which driving forces are involved in the allocation dynamic and an insight on their level of influence into the process.

GIS software by means of spatial analysis tools allows to associate the concept of presence and absence with point futures and generate a point process. This is expressed by a set of distributed points in a terrain. In case of presences, points represent locations of real existing buildings, in case of absences they represent locations were buildings are not existent and so they are generated by a stochastic mechanism avoiding the overlapping with the existent built environment. The location of points is the object of study, which has the objective of understanding its generating mechanism.

The methodology proposes to identify driving forces acting upon landscape affecting rural built transformation and then test the actual existence of a causal relationship between them and the building allocations in the expansion process through a model. Therefore the recognized driving forces responsible for changes represent the hypothesis to be tested. The set up of an hypothesis to be tested through a model recalls approaches adopted for the modelling of population distribution where similarly an initial hypothesis is tested by means of empirical models. Empirical models in fact, statistically relate the geographical distribution of species or communities to their present environment**.**

The model that seeks to be developed is a spatial explicit model because it is by means of the spatial arrangement assumed by new buildings across the landscape that the comprehension of related driving forces is realised. In fact the building site locations is the expression of the action of driving factors in the land suitability sorting process and such condition is synthesized through a spatially explicit modeling approach. Spatial explicit models consider how elements that generate the landscape are changing in space and typically produce maps representing these dynamical patterns. For the specific case study, the outcome map represents, if compared with the actual existent building arrangement, a test for the theoretical assumption formulated.

To test the model developed according to the defined methodology, it is necessary to apply the model in a case study, hence to select a study area as a target, together with a time range that are representative of dynamics under study. The availability of spatial data regarding building locations together with the significance of the landscape motivates the adoption of a study area located in the northern part of Italy within Emilia Romagna Region named New District of Imola (a description of the study area is reported in paragraph III.1). The time range selected for testing the model is the interval between 1975 and 2005 since such period of time is extremely meaningful for transformations occurred on rural built environment.

Two different dataset are applied in modelling building expansion in periurban and rural area, hence two separate calibration processes were carried out. The identification of rural and periurban area refers to the classification adopted for the stratification of the random sampling methodology (see paragraph III.3). In this case study, periurban area is assumed as the output resulting from the spatial difference between urban area existent in 2003 and urban area in 1976 (see fig.9).

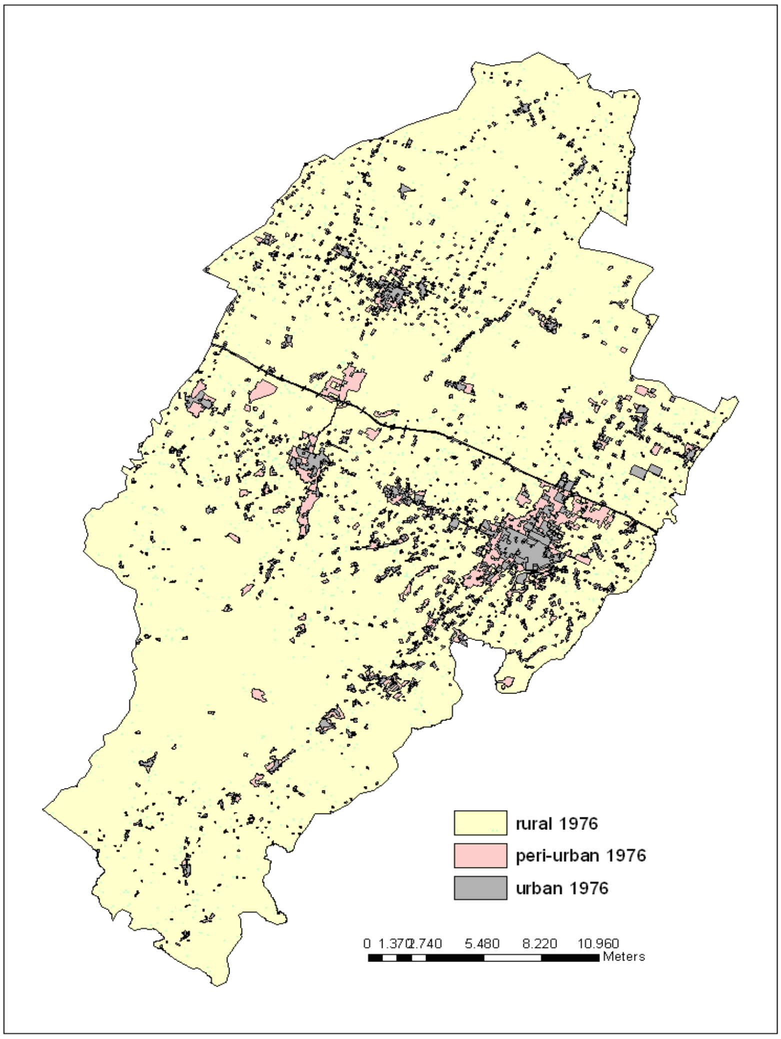


Figure 9: periurban and rural area

The outcome resulting from spatial intersections between sample areas - rural area and sample areas - periurban area, represented the land where spatial data were acquired on which to perform the calibration process.

III.2.2. RESPONSE VARIABLE

The issue at this stage is related to the definition of the best variable able to explicit the building expansion dynamic. Since the objective of the model is the understanding of driving forces having effect on building expansion within rural and periurban area, the dependent variable should be a factor which could explain the variation of buildings over time. In the analysis of building expansion there are two dynamics that seek to be explained: (1) the increasing rate of built up area and (2) locations taken by new buildings. Most models (see Clarke and Herold 2002; Silva and Clarke 1995, White 1993, Verburg et al. 1999), developed with the purpose of analysing the built-up area expansion, have the aim of understanding reasons related to the conversion of different types of land use/land cover and then predict future scenarios. Only few studies were carried out (Aspinall 2003) with the specific intent of focusing on transformations occurred at higher scale in rural built environment. In particular, among different aspects related to built environment transformation in rural area, it seems that there is a general lack of studies dealing with the modeling of rural building expansion. Therefore, it is a specific purpose of this work to provide a modelling methodology for the transformation analysis of building expansion. In particular as explained in the paragraph related to the aim of the study, the interest is focused on the transformation analysis of building allocations.

As previously explained a different combination of factors in various parts of the territory generated favourable or less favourable conditions for the building allocation and the existence of buildings represents the evidence of such optimum. Conversely the absence of buildings expresses a combination of agents which is not suitable for the allocation of buildings. Presence or absence of buildings can be adopted as indicator of such driving conditions and represent the response variable of the model. Sites where building take place are characterized by conditions that acting in a different way within the territory generate perhaps a different evolution of the landscape. It is important hence to set an appropriate variable able to register such existing conditions. Presence or absence of buildings at some site locations represent the expression of these driving factors interaction.

Therefore it is expected that response variable can assumes two values with a discrete distribution along the study area, to register where new edification occurred or not over time. This circumstance can be expressed by a binomial type of variable that assumes the value of 1 for the event corresponding to presence of the building and 0 for the opposite event i.e., absence of buildings.

In fact, for the correct application of the model not only new edifications must be detected and related to their driving factors, but also sites where none building allocation occurred. The expression of variation of building in a quantitative way such as variation of number of buildings or density of building per area, do not represent a suitable approach since would report information solely related to the increasing of built up area not on the distribution that new allocation assumed. It is instead by means of qualitative variable that reasons related to new presence or absence of new building allocation can be understood

GIS software by means of spatial analysis tools allows to associate the concept of presence and absence with point futures generating a point process. In case of presences, points representing buildings are indicated as 1, in case of absences points are generated by a stochastic mechanism avoiding the overlapping with the existent built environment and are indicated as 0. This is a very basic operation that allows to attach environmental information to building spatial locations labelled as points.

The assignment of presence or absence is made on the basis of a diachronic analysis carried out on the same study area. The comparison, between two time steps, allows to locate new buildings. In practice, this is performed at first by overlapping maps of the built environment, represented by polygon features, of the same territory in two different time intervals. Buildings added at the very next time step represent the new building built in favourable sites and therefore indicated as presences. To associate values that driving factors assume in building locations it is necessary to perform a spatial intersection. This operation can not be performed if new buildings, hence those labelled 1, are polygons, because such typology feature doesn’t allows to capture precisely the value of each driving factors existing in correspondence of each building location. The possibility to transform polygon into points overcome this problem.

GIS facilities that handle two-dimensional topological structures do not need a complex of vertices and line segments to define a polygon, but simply a point (a pair of coordinates) called centroid or label. In procedures for reconstruction of the topology, the software performs a research through the entire geographic database to assign to the fields LPOLY (left polygon) and RPOLY (right polygon) of the arc attribute table correct identifiers of individual centroids. For this reason, points and polygons (or points and centroids) share the same list of attributes, which in the case of points by convention, is called Point Attribute Table and in the case of polygon they are abstract items such as centroids of the polygons, Polygonal Table of Attributes. Polygon Attribute Table of contains in any case an area and perimeter. In the Point Attribute Table, these two fields are of course for all records equal to zero. Hence it is possible to transform each polygon representing a building in its respective label point or centroid (fig.10).

In correspondence of point feature locations the response variable assumes value of presence hence labelled with 1.

Since also factors responsible for none building occurrence have to be tested in the model, random points are generated outside building polygons, preventing any overlapping with existing polygon and labelled with 0 (fig.11).

In this way, the variable assumes a binomial distribution which belongs to a type of theoretical probability distribution family and expresses the probability of occurrence of the specific event: building presence or absence.

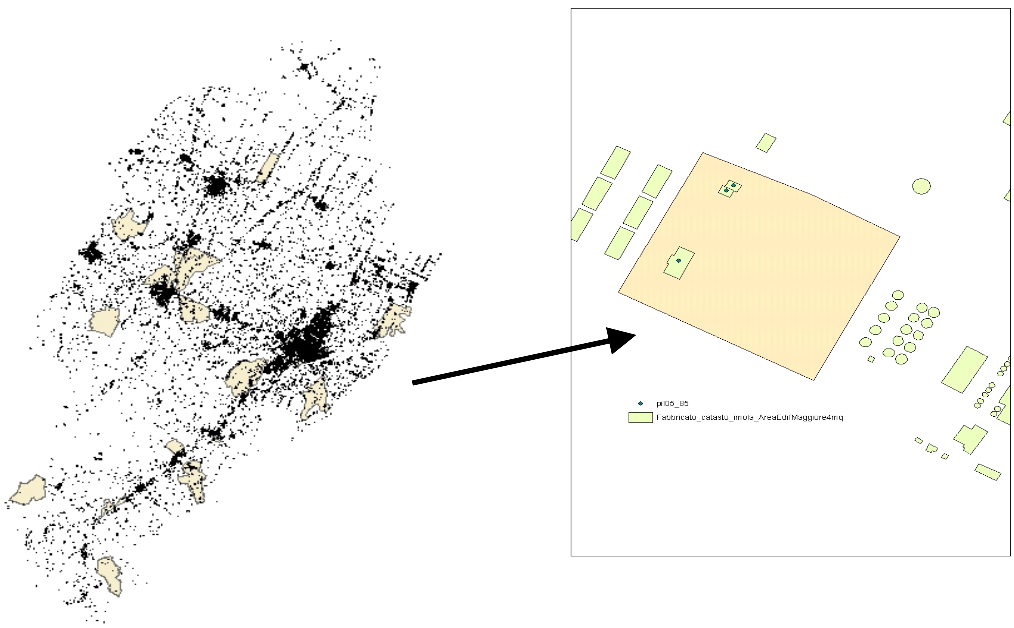


Figure 10 - Generation of label points from building polygons.

Absence of buildings have the function of identifying factors that are not favourable for building allocations. This represents an important component for a correct application of the model. The parameterization of this variable was performed generating randomly a certain number of points equal to the amount of new buildings. Such random points are created within sample areas avoiding any possible overlay with already existing buildings (fig. 11). Their generation required at first the rasterization of sample area and then the use of a specific tool. There are some different tools available for this purpose and in this case was adopted the Hawth's tool.

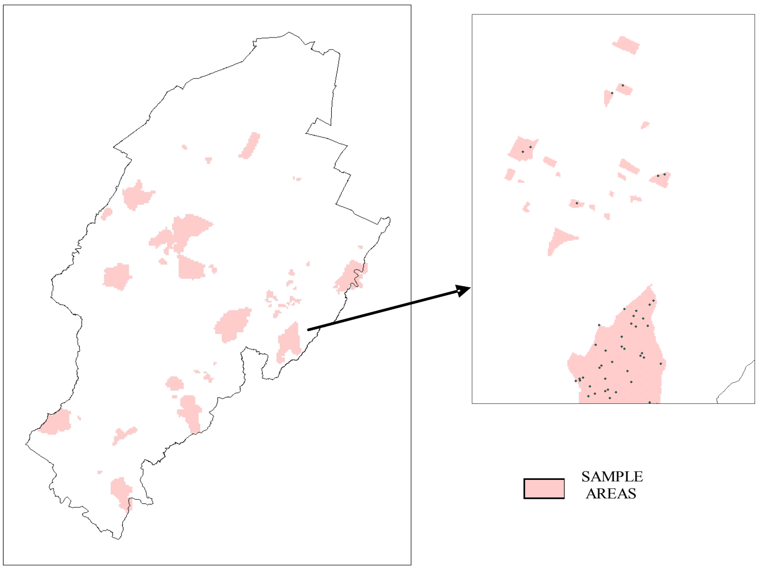


Figure 11 - Generation of random points within sample areas.

III.2.3 DATA COLLECTION: RANDOM STRATIFIED SAMPLING METHODOLOGY

In order to develop a spatial model, an important requirement is the availability of spatial data related to building locations. Mapping building spatial configuration in different time intervals considered for the analysis is essential in order to estimate variation of the built environment. Obviously such requirement involves a large quantities of spatial data referred to built environment in rural area and periurban area. These spatial information are not easily to be collected especially for time interval located in the past where more likely, maps reporting building location for geographical area, are still in paper format and so not useful to carry out spatial data elaboration in GIS environment. Digital format for such type of spatial information in the most part of Italian region became available after 2000. The digital acquisition of spatial data from paper map is extremely time consuming and so demands the application of a sampling methodology in order to extract representative sample area on which to proceed with the acquisition of building spatial locations.

At this purpose a stratified random sampling methodology, resulting from an additional study carried out in cooperation with the research group of the division of Spatial Engineering of DEIAgra (University of Bologna) for a broader research work, is proposed as a sampling data collection model. The referring methodology (see Tassinari, et. al., 2008) represents a fundamental step for the development of the case study. In fact, to test causal relationship between building locations and related acting driving factors it is essential to have spatial information of each new building occurred within a time interval in the study area. The supplying of such parameter, represents a critical phase in the data acquisition process. In particular, since the study focus on transformation occurred in past decades after the second war, spatial information regarding building locations at this time step results as to be the most required. Proceeding backward to seek for information represents a difficult task in data collection process. The specific case of rural built environment analysis relies on data able to provide locations of buildings in the past and in ages not far from the actuality and the conversions of spatial data for buildings distribution which are solely available in paper format into digital vector format is extremely time consuming. On the other hand this step is essential for the comparison of different land arrangements and the evaluation of changes over time. Whereupon the application of a sampling methodology represents a fundamental and compulsory tool for change studies carried out on rural built environment.

The stratified sampling methodology at the end allows to extract representative sample areas of the territory under analysis where to carry out the acquisition and processing of data avoiding to undertake this procedure for the entire study area.
The design of the proposed method for stratified random sampling goes through this following steps:

- Delineation of a suitable sampling framework
- The stratification of the study area according to some variables and criteria for setting population elements
- The extraction of representative sample areas

The stratification methodology was developed on a target study area which is the New District of Imola.
Stratification process involved the identification of variables that mostly affect the development of rural built environment directly and indirectly and on which combination define possible strata. The first stratification variable was the land use, for which 5 classes were defined (this task was carried out in collaboration of GIS department of Emilia Romagna Region).
Land capability connected to elevation were identified as based variable for the stratification because for long time new settlement sites took place and growth where environmental and morphological condition where more favourable to agricultural activities. Hence it seemed logical to aggregated these two variables generating one unique variable named as land suitability for agriculture and forestry production. This is classified in 5 possible categories of suitability to agricultural uses becoming one of the stratification variables (fig.12).

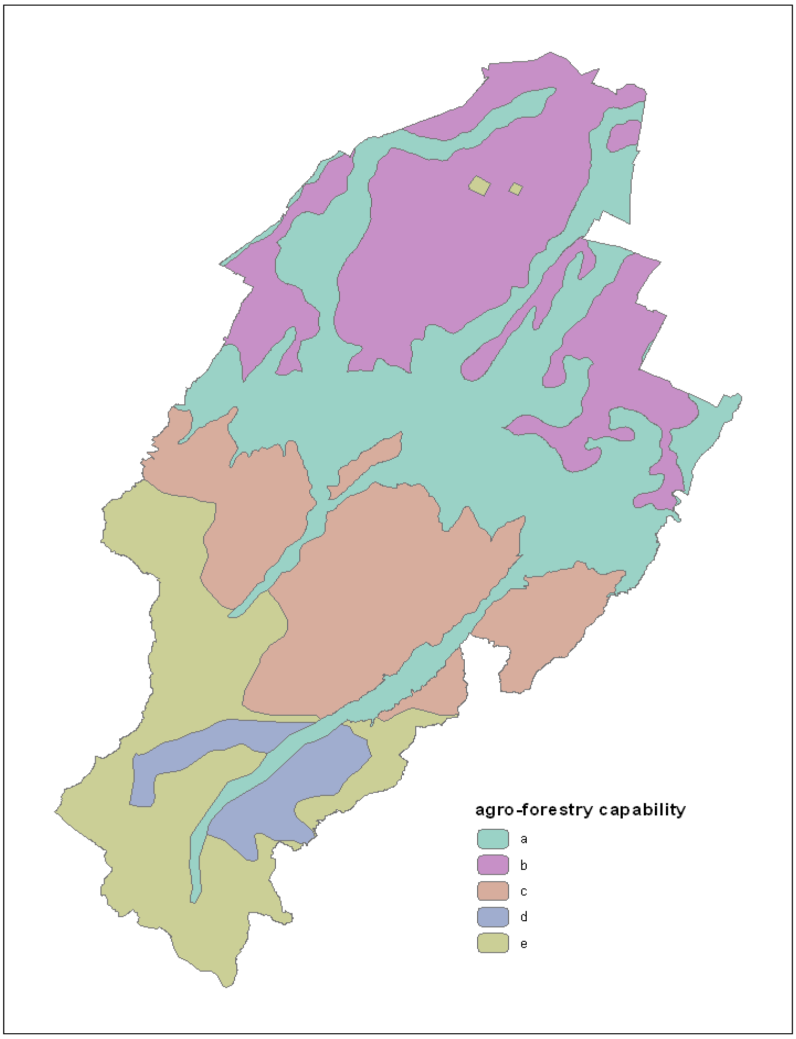


Figure 12 - Agro-forestry capability classified in 5 categories:

Highest level of agricultural land-use suitability (*a*); flat area with vertisoil (*b*); foot-hill and hill area with medium level of agricultural suitability (*c*); lower agricultural suitability of class *c* (*d*); lowest agricultural suitability (*e*).

In area sampling can be chose a sampling frame that subdivide the study area based upon physical boundaries or a lattice built with regular geometrical outline. In the case study proposed, is presented the first option and the area sampling was define according to the most recent Italian population and estate census (Istat, Italian Statistic Board, 2001). Istat provides classification of Emilia Romagna Region by dividing the area into division census polygons of different sizes based upon number of buildings (see fig. 13).

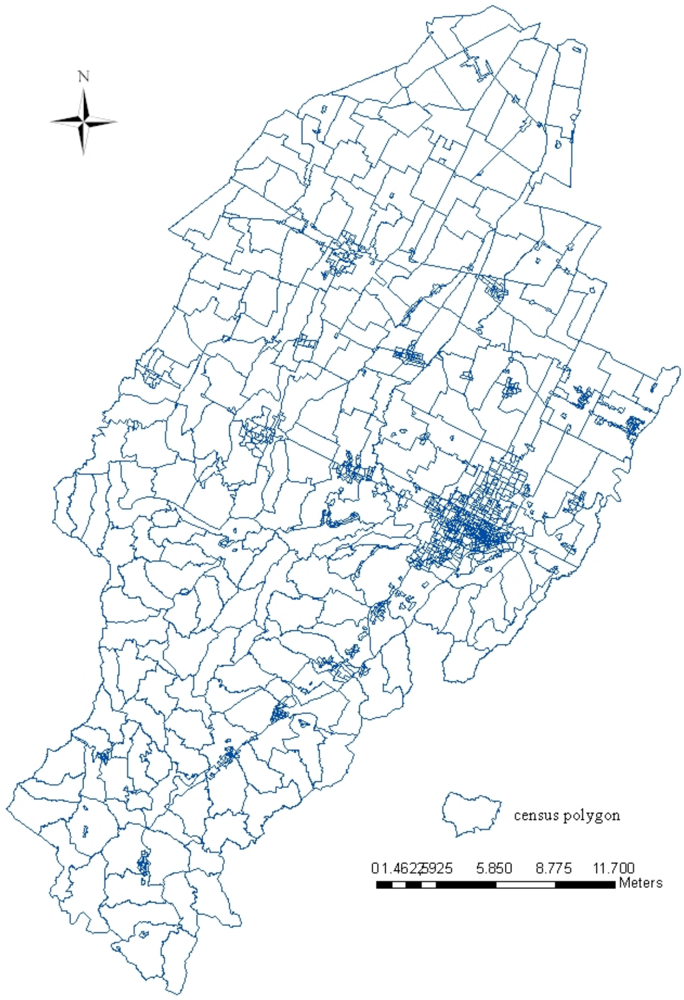


Figure 13 - Istat census polygons

There are four types of census divisions based on the level of buildings aggregation:

- populated centres (highest level of aggregation);
- scatter centres (distances among houses ≤30m);
- productive centres (included in the extra urban area outside of dwellings and towns area where there are at least 200 employed and 10 buildings);
- sprawl houses (none texture of contiguity is observed).

This type of census divisions classification approach is particularly suitable to meet the purpose of defining a sampling frame for the extraction of a representative pilot sample on which carry out studies for rural built environmental dynamics. In fact the extension of census division polygon and hence of sampling frames are sized based on number of buildings, so smaller polygon are in urban area and in its proximity, and larger polygon are in rural area where buildings have more disperse distribution.

The land-use/land-cover represent another variable introduced for the stratification. The data for analysis of land-use/land-cover patterns were acquired from the most recent thematic maps available i.e., those produced by the GIS Department of Emilia-Romagna Region via photo-interpretation of panchromatic orthoimages acquired by the Quickbird satellite in 2003. Such landuse/ land-cover maps, drawn on a scale of 1:25000, present a vector-type geographical structure with a minimum mapping unit of 1.56 ha. The land-use applied as variable for the stratification process was classified in 5 classes (see fig.14):

1. settlements;
2. crop fields and mixed;
3. orchards, greenhouses, garden;
4. silvo-pastoral area, wetlands, area with no vegetation;
5. water bodies and rivers.

The intent of this classification is to suit local characteristics at the greatest level of detail, a key structure referring to the four levels of the Corine Land Cover (European Environment Agency, 2000).

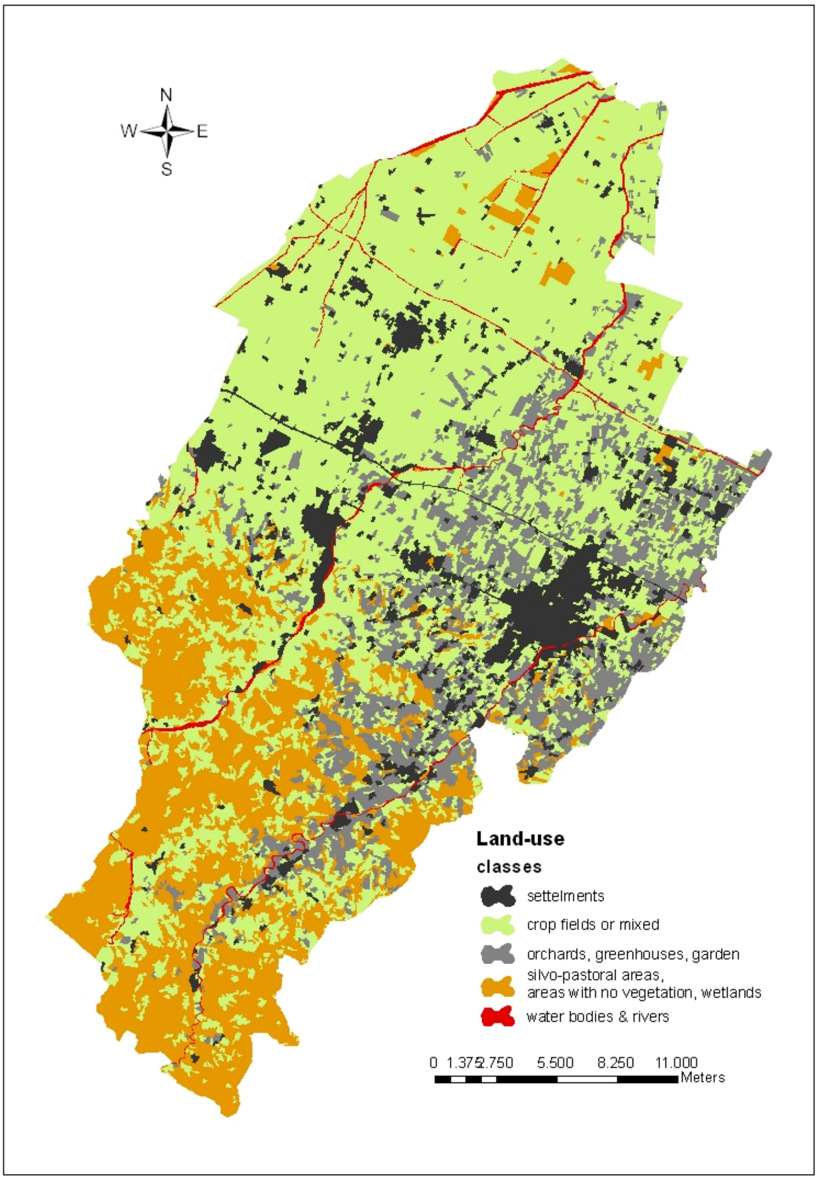


Figure 14 - Land use classification.

Hence the stratification of the Istat census division is performed based upon the combination of the two stratification variables: land-use and land suitability.

The first stage of stratification process consists in identifying all census division occupied by more then 50% of urban area and then subdivide them based on the second stratification variable the land suitability according to a majority attribution criteria. Periurban area is defined as a specific stratum since this area experiences the most marked changes in land use/land cover change with particular transformation dynamics typically of urban fringe.

In this case study periurban area is considered as a particular portion of territory characterized by the presence of typical urban building arrangements such as those assumed by Istat

classification however located in the rural area according to the land use classification and photo-interpretation. Visually this stratum is displayed as the result of the difference between census district polygons classified as populated centre according to Istat classification and census divisions assigned to urban according to the 2003 land- use/land-cover classification.
The classification of extra-urban areas is first carried out identifying the predominant portion of land use/ land cover class in term of extent and then assigning the prevalent land suitability class relative to this. In the specific case study, the necessity of having a small number of strata and the existence of few strata generated by some combination variables, validated the hypothesis of aggregate together certain strata characterized by having the same land use/land cover but different land suitability. Twelve strata were obtained in which each strata is named from the combination of the number of the corresponding land-use/land-cover class assigned and the letter of the corresponding land suitability class. The selection of the sample area can be done using the method of permanent random number (see Carfagna, 2007). More specifically, this involved assigning to each population element in a stratum a real number between zero and one following a uniform distribution and then sorting the elements of the stratum in ascending order according to their randomly assigned number. Following this sequence, a number of elements were then selected from each stratum corresponding to the sample fraction assigned to that stratum by the proportionate allocation criterion. This methodology applied to the study area chosen as target for the calibration of the methodology generated 103 sample areas where was possible to acquire spatial data related to each single building location (see fig.15). Spatial building information within sample area in the past were detected according to a methodology (Tassinari and Torreggiani 2006) that apply a process of backward updating. Maps related to past distribution of rural built system back at 1975 and cadastre at 2005 were compared and buildings that did not yet exist at the previous time step were removed and, if necessary, added where any existing one were subsequently demolished. Spatial information regarding building locations in 1975 within sample areas were collected and mapped in digital format. Consequently, the spatial building expansion was determined performing the spatial overlay of the two maps displaying building distribution in 1975 and in 2005 within representative areas. The new edifications resulting by the difference of the two layers represented the presences of the response variable that was transformed into the value of 1 for the purpose of fitting the GLM.

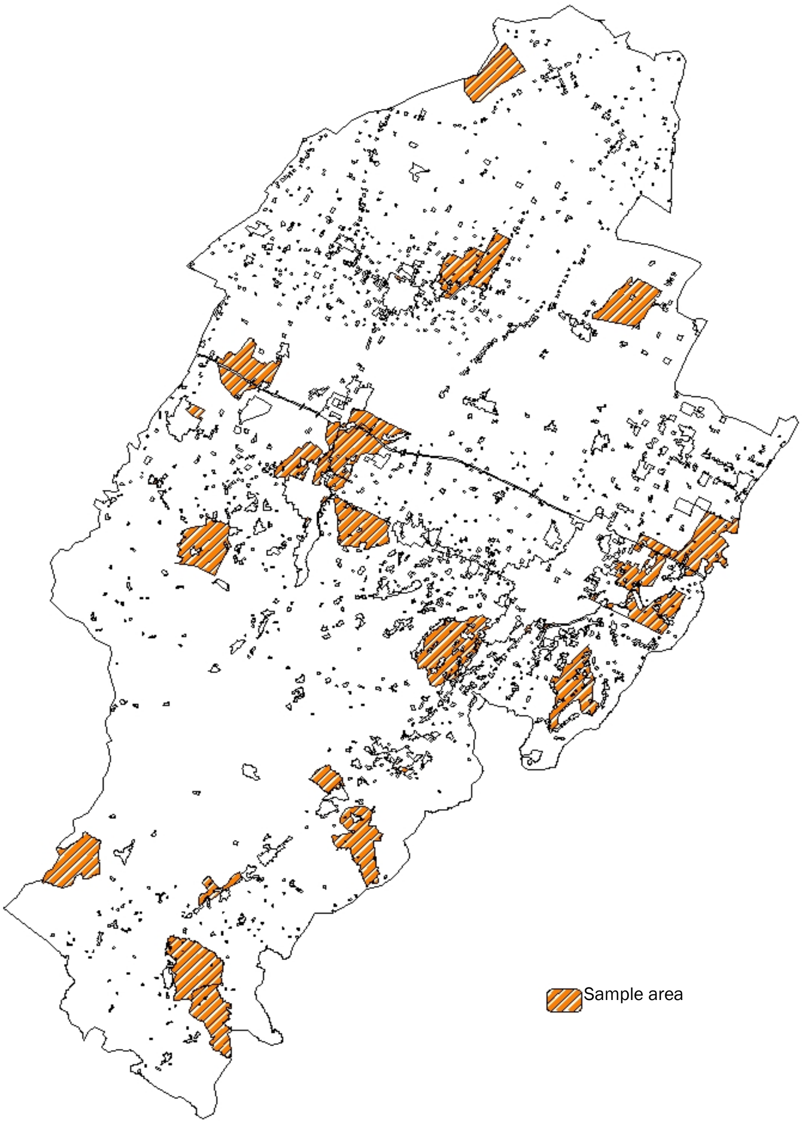


Figure 15 - 103 Sample areas extracted

III.2.4 GENERATION OF DATAFRAME

The spatial building allocation of new buildings resulting from the spatial difference between building distribution in 1975 and in 2005 within sample areas therefore recorded as 1 are labelled as points. Absence of buildings, labelled with points as well, are instead generated randomly still within the sample areas and recorded as 0. These two operations are performed for buildings located in urban and periurban area independently. It is clear that since the spatial processing to determine the allocation of new buildings can be carried out only within sample areas, it is necessary to individuate and extract within these areas, portion of rural and portion of periurban area and on these portions determine spatially the building allocation occurred from 1975-2005 (see fig 16). On the same portion, random points representing absences are generated (see fig 17). The number of random points generated must be the same as points representing new buildings detected in the spatial difference process within sample areas.

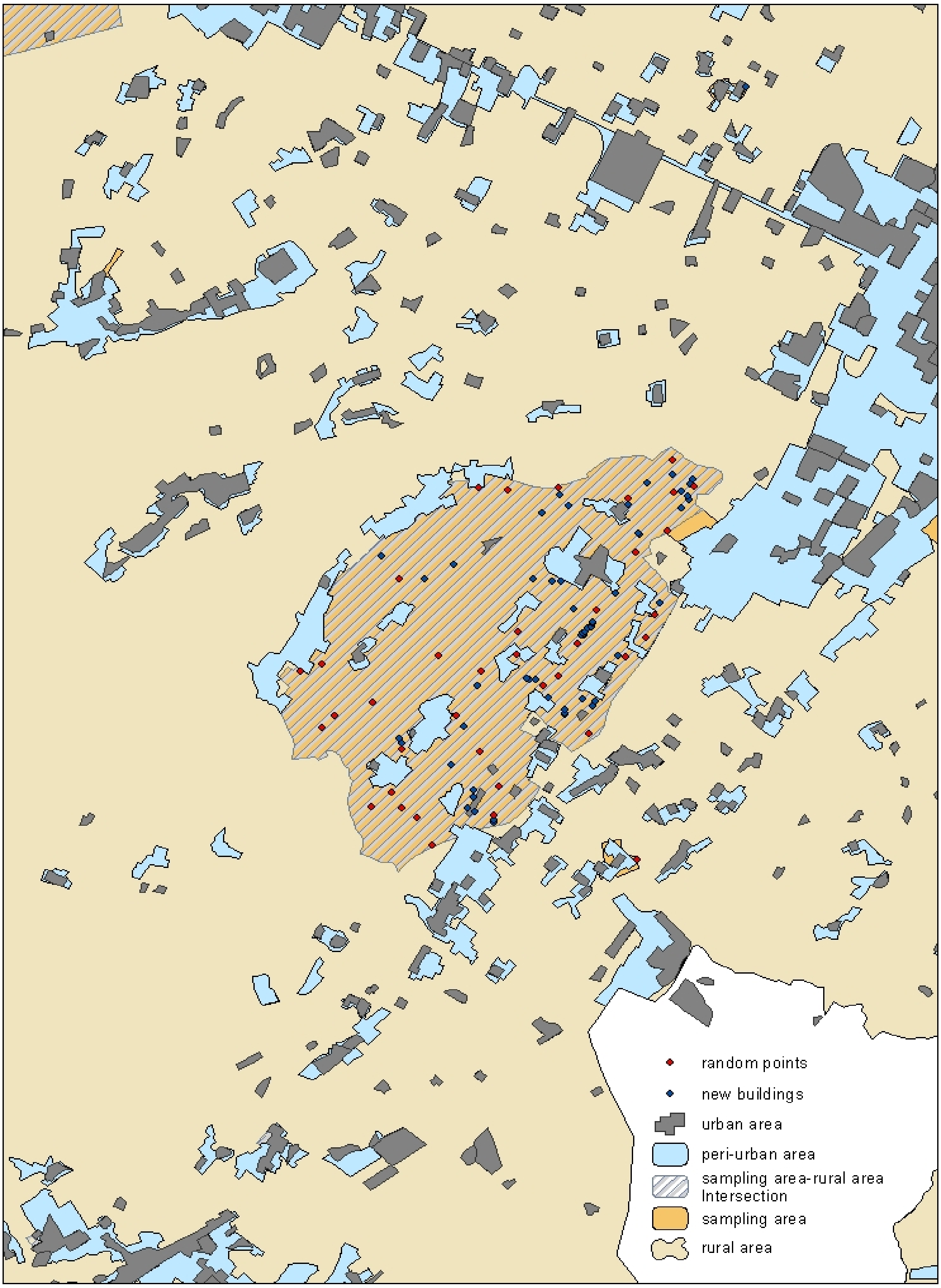


Figure 16 - Presence (new buildings) and absence (random points) of buildings on the layer resulting from the intersection of rural area and sample areas.

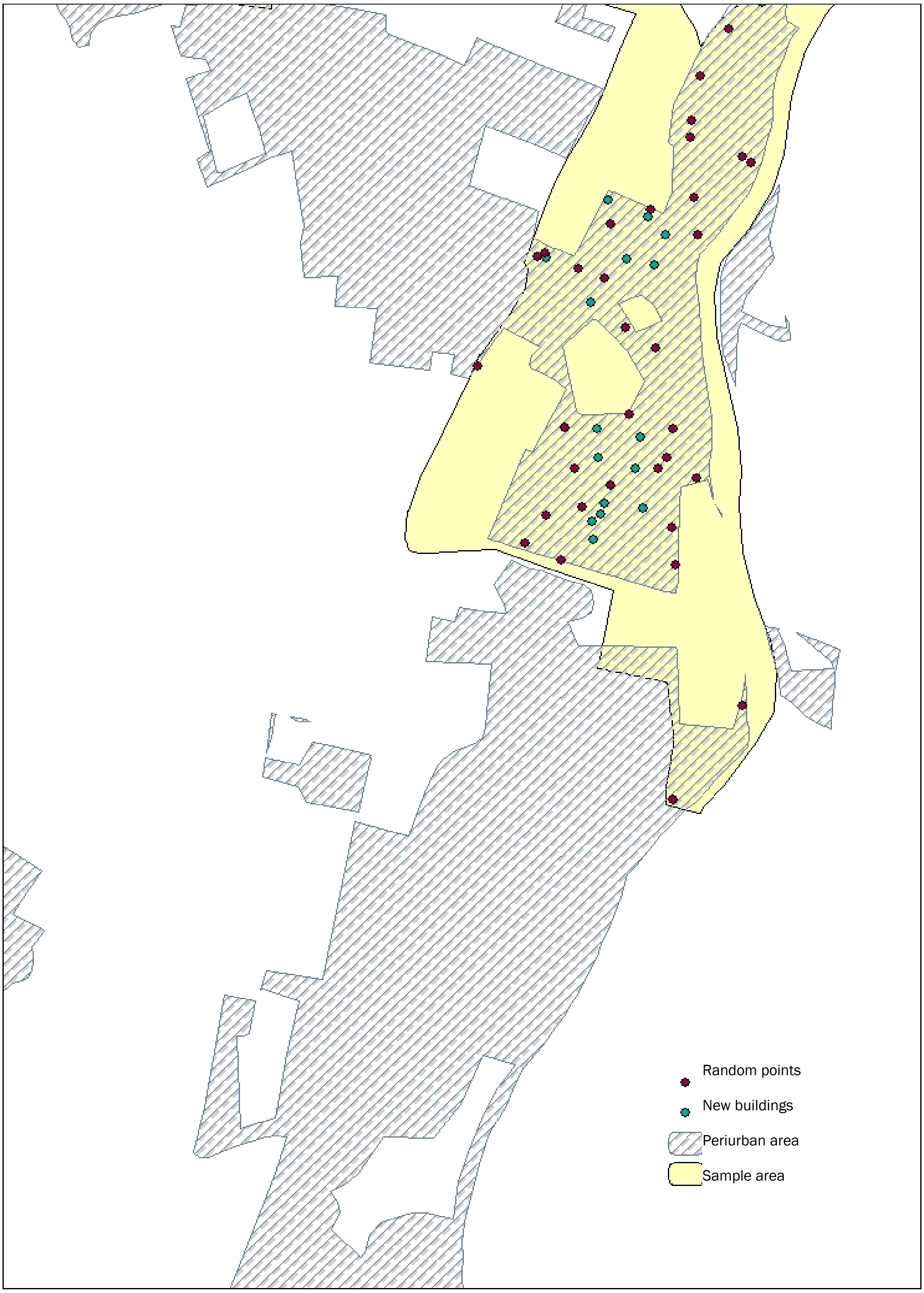


Figure 17 - Presence (new buildings) and absence (random points) of buildings on the layer resulting from the intersection of rural area and sample areas.

The attachment of variable values to points, representing presence/absence of building sites, and the consequently generation of the data frame for the calibration process of rural and periurban model, was carried out by performing an intersection in ArcMap with the Hawth's tool which performs a point-grid overlay (see fig. 18).

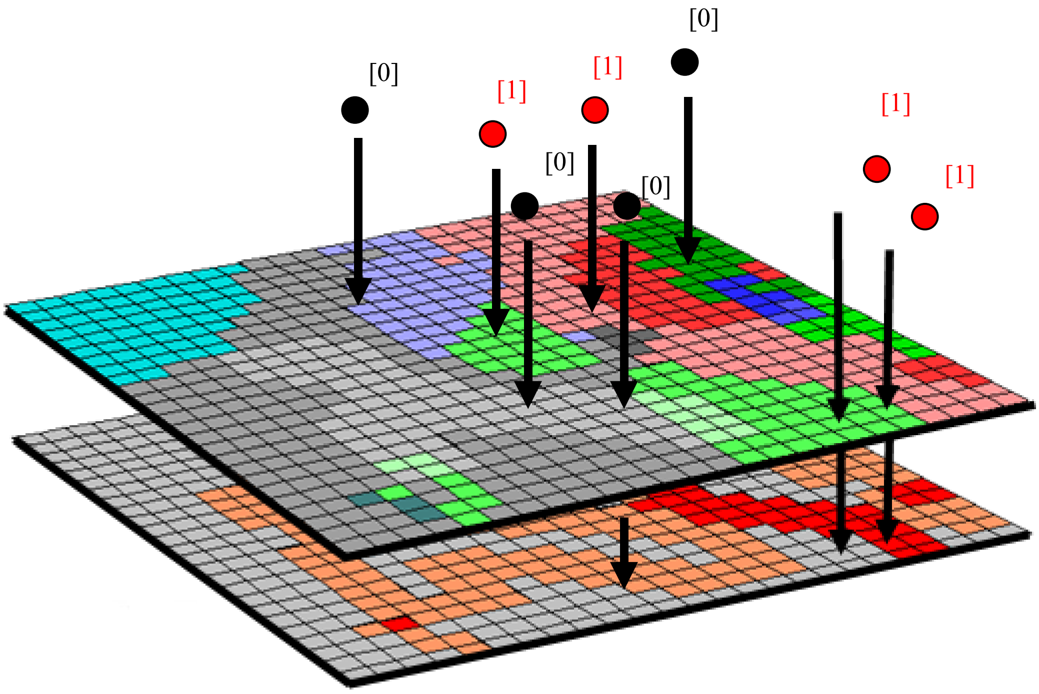


Figure 18 - Hawth's tool point intersection.

The table of contents generated by the overlay process displays for each record the pixel value that explanatory variables take in correspondence of the dropped points that own the presence or absence feature (see table 1).

Table 1 - Table of contents resulting from the intersection of points, that own presence (P) or absence (A) feature, with raster layer of explanatory variables.

P/A	proximity to roads	proximity to periurban	proximity to urban	slope	building density	pop. density	elevation
1	502.49377441400	1215.08227539000	1237.61865234000	11.66663265230	0.78401023150	0.00000000000	543.29693603500
1	484.66482543900	1191.06042480000	1212.20874023000	11.66663265230	0.78401023150	0.00000000000	543.29693603500
1	488.36462402300	1198.85363770000	1223.93835449000	11.75392818450	0.78401023150	0.00000000000	548.17541503900
1	80.00000000000	1606.54907227000	110.67971801800	26.49568748470	0.78401023150	0.00000000000	426.59887695300
1	70.71067810060	1653.98010254000	140.89002990700	28.55367660520	0.78401023150	0.00000000000	416.22967529300
1	36.05551147460	147.05441284200	185.00000000000	11.06765747070	4.99766588211	0.16699999571	160.09582519500
1	10.00000000000	237.53947448700	587.06896972700	25.36139297490	4.99766588211	0.16699999571	180.03361511200
1	793.09521484400	1525.81945801000	1240.25195313000	11.61035633090	0.76076006889	0.00000000000	497.36206054700
...	...	...	...	...	...	...	...
0	778.97369384800	1470.00854492000	1298.47985840000	9.28243732452	0.76076006889	0.00000000000	487.28359985400
0	31.62277603150	2410.85058594000	199.06028747600	7.62764549255	1.82184815407	0.13600000739	384.61099243200
0	20.00000000000	164.46884155300	801.63891601600	6.53141641617	4.99766588211	0.16699999571	137.52951049800
0	14.14213562010	2372.18261719000	241.29856872600	1.42608094215	1.82184815407	0.13600000739	387.39886474600
0	60.00000000000	130.86251831100	848.76379394500	5.33002614975	4.99766588211	0.16699999571	134.95545959500
0	0.00000000000	2349.06884766000	260.86395263700	9.56576442719	1.82184815407	0.13600000739	384.57385253900
0	0.00000000000	2276.92773438000	377.35925293000	15.94096374510	1.82184815407	0.13600000739	351.51116943400
0	58.30952072140	2178.46264648000	385.81082153300	17.78398132320	1.82184815407	0.13600000739	351.06872558600
0	44.72135925290	2160.02319336000	365.85516357400	19.87686538700	1.82184815407	0.13600000739	342.35061645500
0	31.62277603150	1972.23217773000	235.05319213900	24.77745628360	1.82184815407	0.13600000739	309.96017456100
0	70.71067810060	1653.87426758000	203.03939819300	23.81204986570	1.82184815407	0.13600000739	221.89361572300

The table of contents is then exported and converted in a suitable file format to be imported in R where the calibration process is carried out.
was determined performing the spatial overlay of the two maps displaying building distribution in 1975 and in 2005 within representative areas

III.2.5 SELECTION OF EXPLANATORY VARIABLES

One of the most difficult task is to decide which explanatory variables should be entered in the model. The estimation of their coefficients, once they are selected is a straightforward task by means of the calibration analysis. It is in fact the purpose of the study the comprehension of how explanatory variables are related to the rural building allocation.
Explanatory variables selection is extremely linked to the identification of driving forces acting on the landscape; in particular in this specific case the attention focuses on those that are believed to be responsible for building allocations. Driving forces are forces that cause observed change in landscape features and their investigation is the first step in the conceptual model formulation that follows the model selection. Instead, explanatory variable represents a qualitative or quantitative term by means of expressing driving factors so that they can be processed in modelling. Consequently the definition of parameters for the expression of driving forces allows the generation of explanatory variables. There are manly two tasks that must be faced at this stage: the identification of driving forces acting on rural built environment and their measurement for modeling process.
Regarding driving forces identification, literature provides a large amount of resources referring to methods and different analysis frameworks. Anyway it must be recalled that, since driving forces have to be interpreted in a nested scale, because there are always factors behind one that is responsible of changes, some simplifications, depending on the scale of the system under study, must be introduced. Moreover not only the overall of aspects affecting changes often can’t be identified but also it is a difficult task quantify or qualify them by means of indicators.
The expansion of buildings should be predicted on the base of factors that are believed to be driving forces of their expansions. According to Bürgi (2004), at first it is important to define boundaries based on spatial temporal and institutional scale in which the analysis is carried out and investigate on drivers acting within the drafted framework. Every study requires an appropriate scale of investigation set in a framework composed by three axis: time axis

(months, years, decades etc.), spatial axis (lot of parcel, unit of production, landscape element etc.) and institutional axis (group of actors, community, state etc.).

Landscape is a complex system resulting (Brandt et al., 1999) from the interaction of drivers identified in five major groups natural, political, socio-economical, cultural and technological. The investigation of possible driving factors trying to recognized for each of these groups driving factors can also represent a general framework from which to start the analysis.

A carefully reading of landscape elements over time by comparing maps at different time intervals together with a revision of historic past events can be performed to observe the evolution dynamic and identify possible drivers of change and in particular drivers that drugged the allocation of new buildings.

Once driving forces are recognised they need to be measured within the study area hence define parameters that express factors by means of numerical values or categories. It is well known how this step of recognition of suitable parameter to better express driving forces represents a critical aspect into landscape change studies (A.Veldkamp and E.F Lambin 2001, Burgi 2004). However this process represents a necessary effort for the estimation of causal relationship between drivers of changes and the dependent variable through spatial modeling. Sometimes the possibility of quantifying or qualifying drivers indeed represents a constriction into the driving forces process selection. In the specific case of rural built environment transformation, many factors such those referring to planning policies are in many cases not easily to be recognized and measured because of the complexity of planning processes and the difficulty in detecting information regarding past planning actions. Scientific discipline and tradition have caused two distinctly different approaches in the field of land use studies. Researchers in the social sciences have a long tradition of studying individual behaviour at the micro-level, some of them using qualitative approaches (Bilsborrow and Ogondo, 1992; Bingsheng, 1996) and others using the quantitative models of micro-economics and social psychology. Conversely, in the natural sciences, geographers and ecologists have focussed on land-cover and land-use at the macro-scale level, spatially explicated through remote sensing and GIS. Due to the poor connections between spatially explicit land use studies and the social sciences, the land use modellers have difficulties to couple them with the rich stock of social science theory and methodology. This results by the ongoing difficulties within the social sciences to interconnect the micro and macro levels of social organisation (Watson 1978; Coleman, 1990).

To overcome this problem in many cases proxy variables, that are the result of primer driver actions but easier to be measured, are adopted for the modeling process. Usually, also the recognition of landscape elements that work as attractor of the phenomena under study, even if they do not represents primer drivers, is a proficient way to define explanatory variables. For example attractors of building expansion are indeed existence of road networks, facility distributions and service allocations.

One more aspect to be considered in driving forces detection is the subdivision of the study area in periurban and rural area since factors affecting buildings expansion are likely to be very different in the two communities. The study area selected as target for the analysis includes also urban area. Since hypothesis wants to verify the casual relationship influence of selected explanatory variables solely on rural buildings expansion happened in extra urban and periurban area, it is important to test such variables on the corresponding community area. Factors regulating built-up area expansion within urban land are significantly different compared to those existing in rural area which are still diverse from those acting in periurban area. Therefore it is advisable to carry out the investigation of diverse community areas as different systems, in this case periurban and rural, and design an individual model for each one, because not only driving forces type involved can be different, but also their degree of contribution.

EXPLANATORY VARIABLES	Community
Distance from road network	R/P
Distance from urban area	R/P
Distance from periurban area	R
Elevation	R/P
Slope	R/P
Conversion of rural land use types	R/P
Existence of environmental measures	R/P
Relative population density variation	R/P
Variation of relative density buildings	R/P

R = rural
P = periurban

The computation of values assumed by explanatory variables along the study area can be performed through elaborations on acquired spatial data with support of GIS technologies. In particular, the core in the spatial data collecting process related to factors acting upon building allocation is represented by the possibility of carrying out spatial analysis and point raster overlay.

Raster are well suited for representing data that changes continuously across a landscape (surface). They provide an effective method of storing the continuity as a surface and also consistent basis for computing new attribute from existing data. The subdivision of space into regular cells or grids allows through map algebra the generation of maps where the surface assumes continuous values such as elevation values measured from the earth's surface, temperature, concentration, and population density. Cell values can be either positive or negative, integer, or floating point. Integer values are best used to represent categorical (discrete) data, and floating-point values to represent continuous surfaces. Raster also called grid are stored as rows and columns where each location is represented as a cell and at every location is given an object. On the contrary vector model stores the boundaries of objects by means of x,y coordinates, and each object is given a location with its own attribute feature means that the resulting layer is not a continuous surface-data.

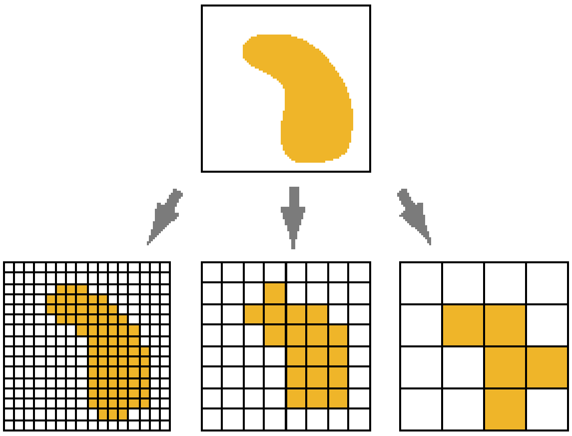


Figure 19- Conversion of vector object, where to each object is given a location, to raster format, where every location is given an object (cell), hence resulting as a continuous surface data.

Continuous spatial data are phenomenon that produce a continuous surface where each location on the surface is based on the inherent characteristics of a location relative to a known fixed point or from an emanating source. They include elevation, slope, distributional map of social economical indices, such as population density, that can be considered as expression of drivers of changes.

For this reason explanatory variable must at first be converted to a raster format in order to be able to extract information from at each point building location.

Once performed this operation from the overlay of points and grids, pixels values can be attached to points and modelling observed data.

It follows the description of explanatory variables identified for the model.

Distance from road network.

Reasons for the variable selection: such variable is an input entered into both rural and periurban model. It is well known how connectivity is important today and the existence of such networks can encourage development of built up area. From the past, new dwellings arose along ancient roads, that still nowadays represent some of the main network axes, generating the first foundation of current towns. The presence of road network, in fact, implies the existence of facilities and services and hence they act as attractors encouraging the development.

Creation of grid: the mapping of this variable is obtained first from the conversion of the road network vector layer into a grid. Cell size is set to 10 meters and the output area extent corresponds to the perimeter of the study area. Consequently the value of 1 is assigned to those pixels corresponding to road network and 0 to all cells laying outside. A grid reporting the gradient of distances from cell having the value of 1 is obtained by calculating the Euclidean distance in archInfo (see fig. 20a).

Distance from urban and periurban area

Reasons for the variable selection: it is believed that urban and consequently periurban centres represent a core in which surrounding the expansion of built-up area mostly take place. The proximity to urban and periurban area benefits of advantages in terms of services and facilities. In many cases rural areas located in the closest surrounding of periurban or urban area are affected by the attraction that these cores represent for other type purposes which differs from the agricultural use and more related to the urban community such as commercial, industrial and residential. This factor along with other reasons are often responsible for the well known urban fringe formation.

Creation of grid: the procedure is similar to the one applied for the generation of the distance from road network grid. Distance from periurban entered only in the periurban model, instead distance from urban take place in both. Grid maps of distance from urban and periurban area were still computed applying the same procedure of calculating the Euclidean distance.

Elevation and slope

Reasons for the variable selection: elevation and slope are considered some of the most important geomorphologic factors that can constrain the generation of suitable or less suitable sites for new building allocation. Where such conditions are more favourable it is expected that more edification occurred. These variables entered into both models.

Creation of grid: elevation is represented by a DEM of a 50 meters cell value, derived from elevation points and contour lines provided by the Cartographic archive of Emilia Romagna Region. Contour lines, with 25 meters intervals, were available only for the southern part of the study area starting from and elevation quote of 50 meters. (see. fig 20b)
Slope is generated from the DEM of the study area through *slope function* in 3D analyst tool of ArcMap and expressed as percentage.

Conversion of rural land use

Reason for the variable selection: the conversion of land use type can have some relations with the building expansion. It is expected that built-up has some causal relation with the transition from a type of land use to another. More specifically areas affected by an intensification of farming activity could discourage the built-up expansion in their surrounding.
Three classes of agricultural land use conversions are considered, to observe the behaviour of building allocation in the transition to one type to the other. The classes of land use considered are: cultivated crops, orchards and forestry. This because they represent by far the most common land-use type within the study area and even if in a large level they can provide the information of how the conversion of land use affected the new building. From a diachronic analysis of land use maps it is possible to observe locations where land use changed along the study area. In particular the possible conversions generated by three classes of land use are 9 combinations as show in the matrix below:

Table 2 - Land use conversion

	cultivated crops	orchards	Forestry
cultivated crops	1	2	3
orchards	4	5	6
forestry	7	8	9

Creation of grid: in this case is a qualitative variable obtained by a diachronic analysis of maps of the same study area. Land use is usually in a vector format with polygons representing classes of land use. By performing the overlay between land use maps that belong to different time intervals and then interrogating the database, it is possible to extract type of conversions occurred over time and locate them. The values reported on tab.2 are assigned to polygons based on the type of conversion occurred. In this way the categorical variable can be divided in 9 levels and converted to raster (see fig.20c).

Existence of environmental measures.

Reason for the variable selection: it is expected that the existence of conservation or preservation measures can represents constrictions to new edifications.

Creation of grid:: this categorical variable is mapped rasterizing the vector map presenting the existence of environmental measures: conservational natural areas and landscape, and parks. The raster output obtained had binary cell values 0/1. Outside of the protected area cell assumes value of 0 and inside the value of 1. (see fig.20f)

Relative population density variation (1991-2001)**.**

Reason for variable selection: population distribution and associated demographic characteristics, e.g. the ratio between urban and rural population, are often considered as the most important factors affecting land use distribution (Bilsborrow et al.1992; Turner et al., 1993 and Heilig, G.K., 1994). Population variation and its distributional trend represent in fact, a proxy driver for social economical factor that can have effects on building expansion. It is expected a causal relationship between them in the direction that an increasing of population should be a trigger for new edification expansion until it is reached a saturation value.

Creation of grid: the index adopted to quantify this variable is the relative variation of population density. The variation of population density was mapped referring to Istat population census data of 1991 and 2001. Such data are reported per division census polygon hence density of population was calculated based on population data per division census polygon. The final raster is resulted from the computation of relative population density variation between 1991-2001 carried out by means of spatial analyst tool in ArcMap. (see Fig.20d)

Relative building density variation (1985-2001).

Reason of variable selection: this quantitative variable gives information on actual building distribution and can provide important explanation on the behaviour of built-up expansion in relation to the neighbourhood.

Creation of grid: Istat provides number of new buildings per year from 1919 to 2001 per division census polygons. The sum of buildings built per year from 1919 to 1985 and the sum of those built per year from 1919 to 2001 provided the total amount of residential buildings existing in 1985 and in 2001. The vector file of Istat division census polygons is then converted to a raster file to calculate the density of buildings in 1981 and in 2001 and then the relative variation of building between 1981 and 2001 with the use of the spatial analyst tool. (see fig.20e)

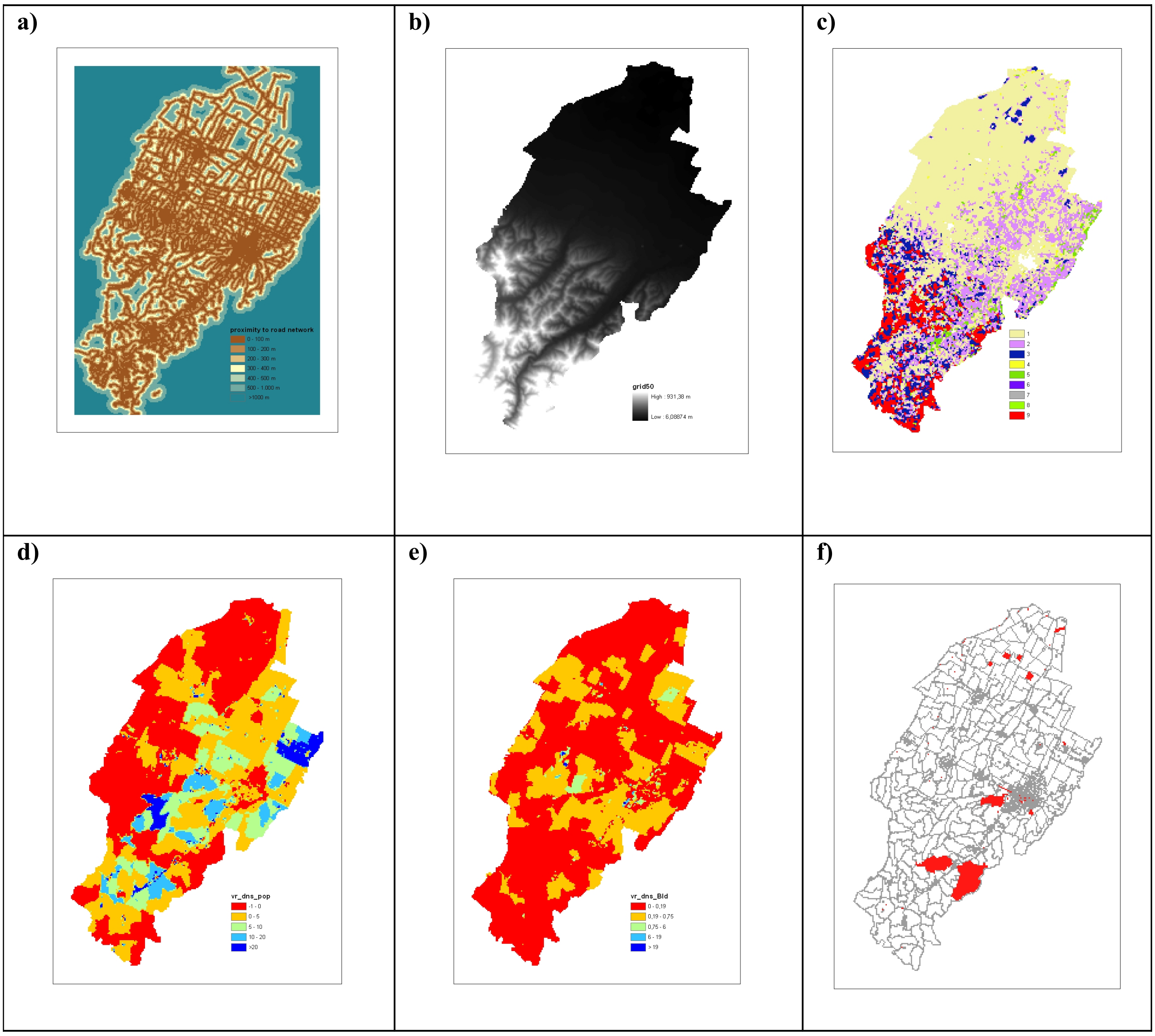


Figure 20 - Distance from road network (a), elevation (b), land-use conversion (c), relative variation of population density (d), relative variation of building density (e), environmental measures (f).

III.2.6. STATISTICAL MODEL FORMULATION

The analysis of existing cause-effect relationship in landscape transformation dynamics represents one of the main research topic in many fields of study with the common main purpose of developing a sustainable planning activity aimed at preserving and protecting landscape resources.

Several ecological indicators have been developed in order to reflect a variety of aspects related to the ecosystem, including biological, chemical and physical aspects, but they are not sufficient in order to understand reasons involved in changing process and to support simulation of future scenario. For such purposes more suitable analysis tools such as modelling techniques, able to understand driving forces acting on landscape transformation and to formulate projections into the future needed to be introduced.

Statistical modeling is used to reach conclusions from data, and there are mainly two goals to be achieved in analyzing data: the extraction of information regarding how the response variable (y) is associated to the input variables (x) and the prediction of response variable on the base of input variables. According to L.Breiman (2001) there are two different approaches toward these goals. One assumes that the data are generated by a given stochastic data model (linear regression, logistic regression etc). The other uses algorithmic models and treats the data mechanism as unknown (decision tree, neural network, etc.). The first is called data modeling culture and the second one algorithm modeling culture (see fig.21,22). More specifically, data modeling approach finds a suitable stochastic data model which is able to detect the type of relationship between response and explanatory variables. The output provides values of the parameters estimated from the data and the model then employed for information and/or prediction. The algorithm approach finds the function that operates on explanatory variables to provide values of the response variable and hence predict y values. The letter methods doesn't provide any information related to how the mechanism works, but an unknown equation that computes the y values.

a)

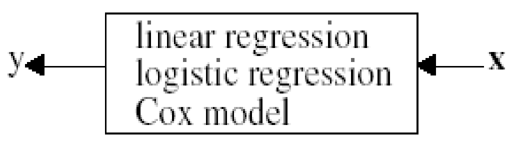


Figure 21 - data modeling culture

b)

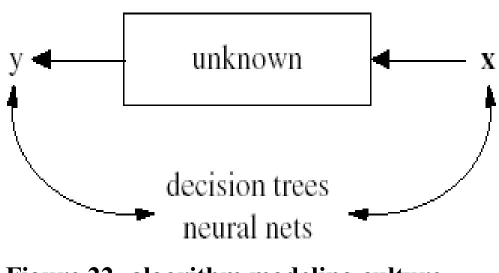


Figure 22- algorithm modeling culture

If the specific purpose is the understanding of reasons responsible for changes it is essential to know the mechanism that generates the dynamic under study i.e., the response variable. At this purpose the data modeling approach, hence the adoption of empirical models provides a mechanism for explanatory variable analysis and for the identification of key driving variables behind the site selection process for new building allocation.

The model formulation represents a phase, into the modeling process, characterized by the choice of a suited algorithm for predicting the response variable and its associated theoretical probability distribution. Regression analysis can be used to make predictions (for example, to provide future data for a time series), statistical inference, hypothesis testing or modeling of relations of dependence. These uses of regression depend on the fact that the starting assumptions are verified. The regression model allows to identify possible autocorrelation between dependent variable and independent variable. In this specific case the type of algorithm put in relation dependent and independent variable providing a measure of how and in which part factors acted on the building expansion.

In linear regression models the vector of observations is taken as the realization of a vector of random variable Y whose components are independent with means μ.

The classic regression model is composed of:

- a systematic (deterministic) component which is function of explicative variables defined as:

  (1) $\mu_i = \Sigma\, x_{ij}\, \beta_i$

β is a vector of unknown parameters object of the inference and x independent variable their association is also called linear predictor η.

- a casual component which is defined assuming independence among observations and the homoscedasticity (i.e., the mean of error ε is equal to 0).

  (2) $Y = \mu + \varepsilon$

  Hence:

  The link between systematic part and stochastic part is an identity:

  (3) $\eta = \mu$

The classical linear regression approach is theoretically valid only when the response variable is normally distributed, the variance is constant and the mean of error ε is equal to 0. However often it is necessary to assess the statistical significance of observed relationships whose data violate the basic assumptions of linear regression.

This occurs, for example, when the distribution of the stochastic component is not Normal as occurs in exponential family distribution.

Nelder and Wedderburn (1972) have proposed a unified approach to a broad class of models that were introduced separately in statistical methodology for dealing with such situations: the theory of Generalized linear models (GLMs).

GLMs constitute a more flexible family of regression models, which allows other distributions for the response variable and non-constant variance functions to be modelled. In the specific case where presence or absence of buildings is assumed as response variable, these two assumptions are violated.
According to Zimmermann and Guisan (2003) most statistical models are specific to the type of response variable and more precisely to its associated theoretical probability distribution. Often more than one technique can be applied appropriately to the same response variable.

A GLM is defined by this following items:

- the casual component: each outcome *y* of the dependent variables, Y, is assumed to be generated from a particular distribution function in the exponential family, a large range of probability distributions that includes the Normal, Binomial and Poisson distributions, among others. The mean,of the distribution depends on the independent variables, *x*, through the link function *(g)* which depends on the type of family distribution (see tab 3).

- the deterministic component: expressed by linear predictor $\beta_j x_j = \eta$. Such component, η , is the quantity which incorporates the information about the independent variables into the model, it is expressed as linear combinations of unknown parameters and β coefficients. The unknown parameters beta are estimated with maximum likelihood. The linear predictor and the link function, describing how the mean of the response and a linear combination of the predictors, are related.

- the link function *(g)*: provides the relationship between the linear predictor and the expected valued of the distribution function. There are many commonly used link functions, and their choice can be somewhat arbitrary. It can be convenient to match the domain of the link function to the range of the distribution function's mean.

In this specific case the type of algorithm put in relation dependent and independent variable providing a measure of how and in which part factors worked on the building expansion.

Table 3 – link functions and inverse link functions for some of the probability distributions.

Family	**Method**	**Link Function**	**Inverse-Link Function**
Gaussian	Identity	$\eta = y$	$y = \eta$
Binomial	Logit	$\eta = ln(y/(1-y))$	$y = [exp^{(\eta)}/(1+exp^{(\eta)})]$
Poisson	Log	$\eta = ln\ (y)$	$y = exp^{(\eta)}$

In the specific case the response variable is expresses by binary data on the presence-absence (0/1) of buildings and a good approximation is to use a Binomial variable which has a Binomial distribution hence belongs to the exponential family.

Such dependent variable is presence/absence, neither of which meets the assumption of an unbounded dependent variable. Presence/absence is either 0 or 1, with no intermediate values possible and no values less than 0 or greater than 1. In this specific case because response variable belongs to the binomial family the logistic regression is applied.

In statistics, logistic regression is a model used for prediction of the probability of occurrence of an event by fitting data to a logistic curve. It is a generalized linear model used for binomial regression. It makes use of several predictor variables that may be either numerical or categorical. In logistic regression, the dependent variable is a *logit*, i.e, the natural log of the odds. The log of the odds overcomes the fact that the dependent variable is not in the same scale of independent variables.

(4) $g = ln(p/(1-p)) = \eta$

g = logit
p = probability of the event
$p/(1-p)$ = odds
ln = natural logarithm
η = linear predictor

Then the odds have to be converted to a simple probability of occurrence through the exponential function to return values in the scale of probability of occurrence:

(5) $p = e^{\eta}/\ (1+ e^{\eta})$

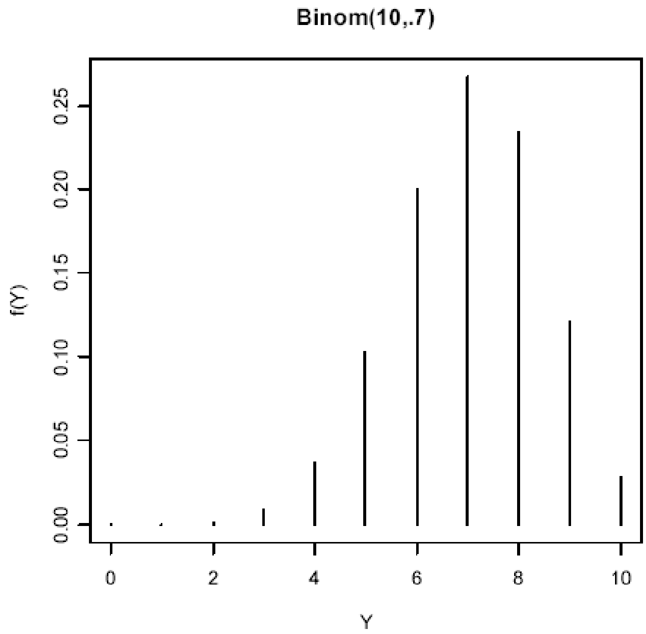
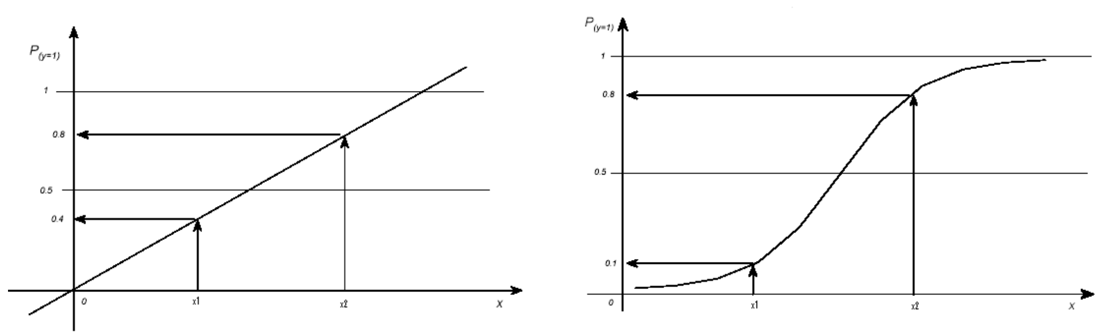


Figure 23- Binomial distribution.



a) **b)**

Figure 24 - a) linear regression and b) logistic regression. When the dependent variable is dichotomic the link function allows that values included into the structural component are within the interval 0-1 so that also expected value will respect the same interval

III.2.7 MODEL CALIBRATION

Ryekiel (1996) defines calibration as "the estimation and adjustment of model parameters and constants to improve the agreement between model output and a data set". Hence model calibration is an adjustment of the mathematical model that was selected for the dataset. According to Zimmerman and Guisan such definition should be enlarged to encompass the more global phase of model construction, which includes also the selection of explanatory variables.

Models are often calibrated and verified using past history on the grounds that the future will repeat the past.

The fit of most models is characterized by a measure of the variance reduction or deviance reduction such as in case of the maximum-likelihood estimation performed by GLM. In GLMs, the model is optimized through deviance reduction, D^2 (see formula 4), representing the equivalent of R^2 in Least Square (LS) models.

(6) D^2= (Null deviance – Residual deviance) / Null deviance

Null deviance = the deviance of the model with the intercept;

Residual deviance = is the deviance that remains unexplained by the model after all final variables have been included. A perfect model has no residual value so its D^2 takes the value.

In GLMS the calibration process is performed by applying the proper link function. The link function provides the relationship between the linear predictors and the mean of the distribution function. There are several canonical used link functions (table 3), and their choice depends mostly on the response variable family distribution, in R this association happened by default (formula 3). Therefore a transformation of y, $g(y)$, not y itself is related to the predictors:

(7) $g(y) = \beta_0 + \beta_1 x_1 + \ldots + \beta_k x_k$

β_0 = intercept

β_k = estimated coefficient of the variable x_k. It is the amount by which the correspondent x_k variable is weighted

x_k = is the value of the variable (explanatory variable) and it is often called predictor.

Example of GLM formulation in the R syntax:

(8) *glm (y ~ x1 + x2 + x3, family = binomial, data = fb)*

Running the GLM the following equation terms are provided as outcome:

- Intercept: is a constant value that tells the expected value when the value of all explanatory variables is 0.
- Beta estimated coefficient: is the amount by which the x_k variable is multiplied and weighted and so it is often called the b coefficient or b weighted. It indicates how much change in y is associated with a change in the variables x_k. It must be recalled that beta values are in the scale of the linear predictor (after transformation by the link function), not in the scale of the response variable (otherwise the values would be between 0 and 1).
- Significant standard error: the standard deviation of the differences between the actual values of the dependent variables (results) and the predicted values.
- z-values: it is often desirable transform the raw scores to standardized scores so that their magnitude can be easily assessed. At this purpose it is provided the z-score in the analysis output which is determined dividing the parameter estimated by the standard error. It indicates how many standard deviation units the value lies from the mean of the distribution.
- p-values: represents the probability that the variation between conditions may have occurred by chance, so variables with smaller p-values are less likely to have occurred by chance, and are conversely more likely to be related to some conditions. It gives indication whether the variable has statistically significant predictive capability in the presence of the other predictors

The goal of logistic regression is to correctly predict the category of outcome for individual cases using the most parsimonious model. To accomplish this goal, a model is created that includes all predictor variables that are useful in predicting the response variable and one of the most difficult task is to decide which explanatory variables, or combination of variables brings significant contribution to the variability of the model. According to Harrel et al. (1966) the number of explanatory variables entered must be reduced to a reasonable number. Several different options are available during model creation. Variables can be entered into the model in the order specified by the researcher or logistic regression can test the fit of the model after each coefficient is added or deleted, called stepwise regression. The selection of predictor can be made arbitrarily (which is not recommend), automatically (e.g. by stepwise procedure) or following some factor behaviour rules. It is well known that none of them is guaranteed to select the best k predictors. Scientific literature can just provides some hints.

In this specific case was applied an automatic selection process by stepwise procedure. Stepwise selection is a frequently applied method to achieve variable reduction that allows dropping or adding variables at the various steps of the model and test in each step the amount of deviance reduction so to understand by means of AIC (Akaike's information Criteria) which variables add more change. AIC is a common statistical measure for model comparison. Stepwise is usually applied in a forward, backward or both ways. Forward selection starts with inclusion of the most significant candidate covariable in a regression model. Backward selection starts with elimination of the least significant one from a regression model that includes all covariables (a full model). The overall fit of the GLM of models generated by stepwise procedure, by adding or deleting variables, are compared by means of AIC measure which establish the best model.

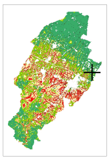
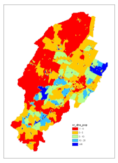
The calibration allows the determination and adjustment of model parameters. Such parameters, also called *beta* (β) coefficients, evaluate the entity of explanatory variable (x) contributions to the determination of the response variable (y). The sign of the parameter has an its own importance. Negative values express a negative contribution of the explanatory variable to the response variable event. The interpretation of the parameter contribution must be done carefully. In fact, since GLM adopted a link function to transform the response variable, that belongs to probability distribution family, into linear scale like predictors (i.e. explanatory variables), estimated parameters are not into the scale of the response variable. Prediction made with this formula will be in the scale of the Linear Predictor , not in the scale of the response variable. Thus, these predictions needs to be inverse-transformed back to the scale of the response variable, using the inverse of the link function. In the case of a Binomial GLM, the logit link function is reported in table (3)

It is $g(y)$ not y which is related to the predictors in the GLM formula. So, considering linear predictor as the untransformed predictions, they can be converted to the scale of the response variable using the inverse function.

In a GLM can be also introduced qualitative explanatory variables of dichotomical type through dummy variables (D), such as for environmental measures or variables of convenience . These are indicator variables that assume a value of 1 if the feature is owned and qualitative value 0 if it is not owned.

(9) $g(y) = \beta_0 + \beta_1 D + \beta_k x_k$

The GLM is then implemented in ArcMap. In order to do this, it is necessary to reproduce the formula where each fitted coefficient multiplies its corresponding variable .

(10) $g(y) =$ intercept $+ \ \beta_1$  $+ \ \beta_2$  $+ \ \beta_n$ 

To return the expected value (y), i.e. the probability of building occurrences distribution map, the linear predictors must be transformed into the scale of the response value by applying the inverse of the Logit link (see tab.3)

III.2.8 *EVALUATION OF THE MODEL*

The accuracy of the model represents a test for the assumption theory formulated. A common approach in the evaluation process is to divide the data into a calibration set and a verification set, using the former to determine the best values of any unknown parameters and using the latter to verify the model's predictions. If observation number is not sufficient to split the dataset in two parts leaving a portion of it out for the verification process. In fact there are types of validation tests that can be performed on the same dataset used for the calibration process.

However in this particular case, since the model is developed on the base of a dataset acquired by sample areas (see paragraph III.2.3), another method should be defined in order to test the probability of building occurrence as estimated by the model against the effective allocation of building in 2005.

The correspondence between probability of building occurrences and actual building features is tested by performing the mean of those probability cell values that are located where actual buildings lie. Probability map have a range of values ranked from 0 to 1. The highest probability of having building allocation is expressed by 1 and vice versa 0 expresses the lowest probability.

From the overlapping of point distribution map of buildings in 2005 with the probability distribution map estimated by the model, if the model worked properly there should be a large amount of cells having value of 1 in correspondence of point features. Collecting these cells and performing the mean, it is possible to establish in which measure the model was able to detected real building distribution on the base of recognised driving forces. Higher is the mean score obtained by the computation, greater is the power of the model.

For the proper work of the method it should be tested not only the spatial correspondence between pixels with high probability and the built environment actually existent, but also the correspondence between low probability values and the absence of the built environment. In this way it can be verified not only if the model is able to estimate the building occurrence where there is actually built environment, but also where buildings do not exist.
The test is carried out individually for urban and periurban area and separately inside boarders of each municipality of the New District of Imola. This mainly allows to evaluate more in detail where driving forces acted differently in some portions of land.
Once again with the Hawth's tool is performed the intersection between point features representing buildings in 2005 and cell values of estimated probabilities. The same procedure is followed for not existing buildings, represented by points generated randomly. As a result, a new field is added into the attribute table reporting point features of "existing buildings in 2005" and into the attribute table reporting point features of "not existing buildings". Thanks to the point intersection, the new field, registers values of the corresponding pixels in the overlapped probability map. Based on these collected values it is then possible to compute the mean for each municipality and evaluate the goodness of the model in detecting absence and presence of buildings.
The accuracy of the model referred to the dataset acquired from sample areas can be tested by performing an evaluation test. Several tests are available in literature, the selection in many case depend on the amount of dataset available for the calibration of the model. When no independent dataset is available, thus there is only one dataset available at the beginning of the study as in this case, basically two methods can be adopted cross validation or bootstrap.
In cross-validation, the data, the same used for the calibration process, are split into K equal size partitions (or nearly equal). The model is fitted with all partitions but one, and the model is then evaluated on the observations in the left out partition. This operation is repeated K times, one for each left out partition. In this way, the data used for evaluation are pseudo-independent, as they were never used for fitting the model. In bootstrap, the data are randomly resample with replacement (i.e. a same observation can be sampled many times). A model is fitted on each sample and the error can be estimated from the empirical distribution of model coefficients across all samples. Hence, as commonly used in this field, the bootstrap is more an approach to calculate robust standard errors. Consequently cross-validation test was run on R as evaluation test of the designed model.
With a sufficient amount of data the model would be developed solely counting on part of the dataset. Hence the expected values determined by the model are compared with the original

values and tested to the original value. Since no surplus of data is available, every time the model is fitted with all observation but one, then predicted values obtained from the model are compared with original data. Several evaluation metrics are available to perform and express such comparison. One such measure is the AUC, the area under the curve of a ROC (Receiver Operating Characteristic)-plot.

It provides two measures derived from the 2x2 contingency table - sensitivity and specificity – that are calculated for all thresholds between 0 and 1, and their values are plotted in a graph of sensitivity (y-axis) versus [1 – specificity] (x-axis). The points corresponding to the successive thresholds are aligned to form a curve. With a perfect agreement, this curve would follow the y-axis on the left part of the graph and the x-axis on the top of the graph, thus passing by the point of sensitivity = 1 and (1- specificity) = 0. Taking the area under such curve would provide the value 1 (as both axes range between 0 and 1). On the contrary, a prediction not different than obtained by chance alone would yield a 1:1 line, below which the area would be 0.5. Hence, the scale is between 0.5 (random prediction) and 1 (perfect prediction). Usually, and according to Swets (1988), AUC < 0.7 means poor predictions, 0.7<AUC<0.9 means useful predictions, and AUC>0.9 means good to excellent predictions.

IV RESULTS

This chapter provides the results of the proposed study that consists in applying the model as defined by following the methodology developed so far, to a case study. The calibration phase, carried out by entering the selected explanatory variables, and the evaluation phase of the model are both shown and performed following approaches defined in the methodology .
The modeling process is carried out for a study area (described in paragraph III.1) where both rural and periurban communities are represented and, for reasons that were previously explained, the two areas are processed separately. Therefore the calibration process is carried out on two dataset, one refers to rural and the other to periurban area: for both areas the time range considered for the calibration process within 1975-2005.
The calibration process allows to estimate the causal relationship between dependent and explanatory variables.
Parameter estimation an objective mathematical operations are available in many statistical packages (SAS, S-PLUS, SYSTAT). In this case, calibration process is carried out in R software with implementation of GIS software for the generation of the probability map..

IV.1. *MODEL CALIBRATION OF RURAL AREA*

The following explanatory variables, identified as possible factors of change, as explained in paragraph III.2.4, are entered as equation terms to fit GLM:

1. Distance from periurban area
2. Distance from urban centres
3. Relative variation of residential building density (1919-2005)
4. Relative variation of population density (1991-2001)
5. Elevation
6. Slope
7. Distance from road network
8. Agricultural land use conversion

To fit the model some transformations of equation terms need to be attempted. In particular, since some explanatory variables show a positive skewness in the response distribution (i.e. the most part

of the data set is located on the left side and few values distributed on the right side) in order to reduced it, logarithmic and quadratic of the logarithmic transformation are attempted for each independent variable. In fact, it is tendency of the logarithm function to compress higher values in the data set and stretches out smaller values.

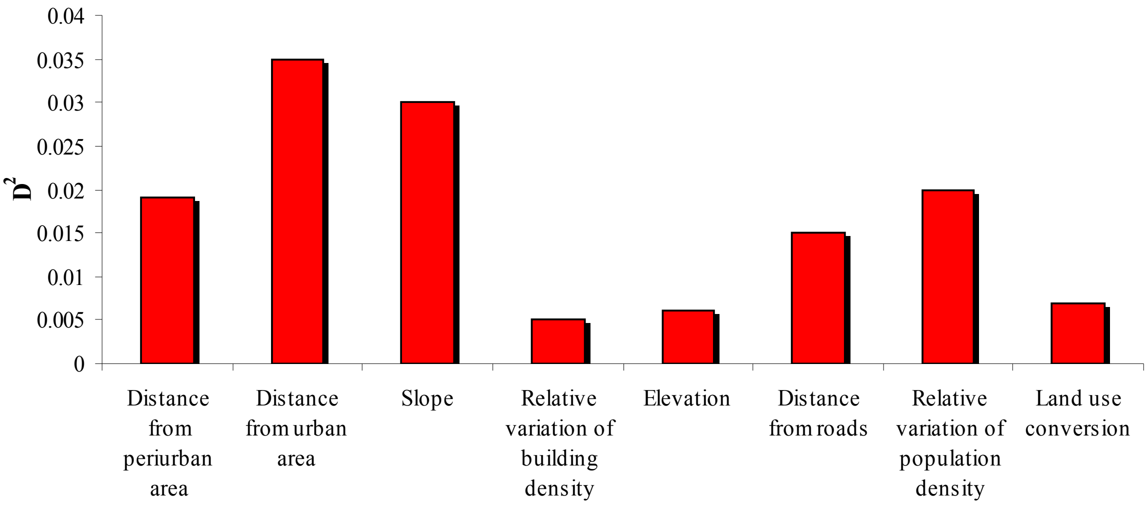


Figure 25 - D^2 values of each explanatory variable.

In addition to log transformation, to achieve an higher D^2 value, hence a better fitting of the model, quadratic and quadratic of the log transformation were also performed. The criteria followed for the selection of the most suitable adjustment, among those attempted, is to adopt the one that combine the best D^2 value achievement with the most simple transformation.

Consequently to the adjustment process, new equation terms are entered in the calibration process providing the improvement of D^2 values of the following variables (see fig.26):

1. Distance from periurban area: 0.019 → 0.03
2. Distance from urban area: 0.035→0.036
3. slope: 0.03 → 0.05
4. relative variation of building density: 0.005 → 0.053
5. elevation: 0.006 → 0.09
6. Distance from road: 0.015 → 0.042
7. relative variation of population density: 0.02 → 0.059

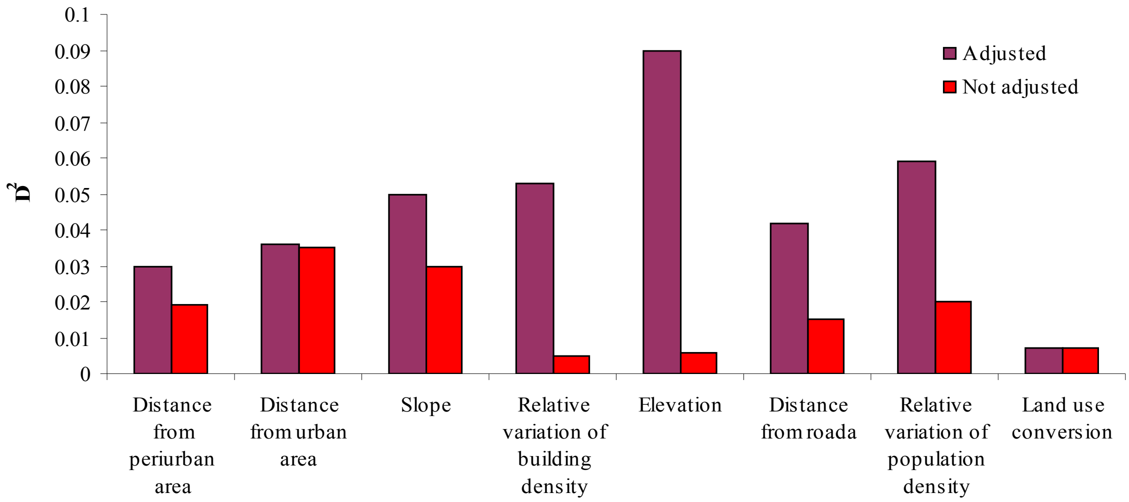


Figure 26 - The increasing of D2 values after the adjustment transformations.

The output from the running of the model in R is represented by a table (see table 4) which provides the following items: the intercept values, estimated beta values, standard error of estimated Z-score and the p-value (the meaning of each of these measures is explained in paragraph III.2.7).

Table 4 – Output table of model calibration.

Coefficients:	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	4.75E+01	8.91E+00	5.33	9.82E-08 ***
log(proximity to perurban area + 100)	1.12E-04	6.70E-04	0.167	0.867424 **
log(proximity to perurban area + 100)^2	-1.67E-07	3.05E-07	0.548	0.583618 **
distance to urban area	5.90E-04	9.95E-04	0.593	0.553504
distance to urban	-6.99E-07	7.75E-07	-0.902	0.366948
log(slope + 1)	3.46E-01	4.64E-01	0.746	0.455709
I(log(fb$slope + 1)^2)	-4.39E-02	1.14E-01	-0.386	0.699686
log(rel.var.building density + 1)	3.93E+00	2.03E+00	1.932	0.053309 .
log(rel.var.building density + 1)^2	-5.30E+00	3.13E+00	-1.694	0.090317 .
log(elevation+ 1)	6.95E+00	1.84E+00	3.783	0.000155 ***
log(elevation + 1)^2	-7.79E-01	2.09E-01	-3.725	0.000195 ***
conversion of land use 1	-9.51E-01	5.07E-01	-1.874	0.060873 .
conversion of land use 2	2.77E-01	3.48E-01	0.794	0.427188
conversion of land use 3	2.53E-01	4.98E-01	0.509	0.61066
conversion of land use 4	8.23E-03	3.85E-01	0.021	0.982951
conversion of land use 5	2.04E-01	3.22E-01	0.634	0.526319
conversion of land use 6	-2.35E-01	1.94E+00	-0.121	0.903871
conversion of land use 7	4.60E-01	6.00E-01	0.766	0.443474
conversion of land use 8	2.94E-01	1.50E+00	0.197	0.843924
conversion of land use 9	-2.60E-01	4.11E-01	-0.632	0.527102
rel.var. pop. density	7.79E-02	4.54E-02	1.716	0.086113 .
(rel.var. pop. density)^2	-3.14E-03	1.45E-03	-2.173	0.029774 *
log(proximity to roads + 100)	-2.22E+01	3.33E+00	-6.659	2.76E-11 ***
log(proximity to roads + 100)^2	1.91E+00	2.96E-01	6.442	1.18E-10 ***

Legend:
rel.var.building density : relative variation of building density
rel.var. pop. density: relative variation of population density
log: logarithm
^2: quadratic function

It is important to specify that all estimated value are in the scale of the linear predictor (after transformation operated by the link function, see paragraph III.2.7), not in the scale of the response variable, otherwise the values would be ranged between 0-1 and they would express probability of occurrence.

As previously explained, to enhance the highest accuracy in the fitting of the model, the number of explanatory variables entered must be reduced to a reasonable number. Therefore it is applied an automatic selection process by means of stepwise procedure. Automatic stepwise selection is easily performed in R in the forward, backward or both modalities. Forward selection starts with inclusion of the most significant candidate covariable in a regression model. Backward selection starts with elimination of the least significant one from a regression model that includes all covariables (also called full model). The option "*both*", which is the option adopted in this specific case, allows to perform both the modalities and so to have a more accurate selection of significant explanatory variables.

It is important to clarify that even if a variable has a low estimated beta coefficient that doesn't necessarily means that such variable doesn't provide a high degree of variability to the model.

The amount of deviance reduction is expressed by AIC (Akaike's Information Criteria) measure. The overall fit of the GLM generated by stepwise procedure, by adding or deleting variables, are compared by means of AIC measure which provides a term by means of establishing the best model.

The stepwise selection procedure recognises the following equation terms as significant to enter in the model:

Table 5 - Output from stepwise selection: (-) selected variables; (+) discharged variables.

- log(distance from periurban area + 100)
- log(distance from periurban area + 100)^2
- log(rel.var.building density + 1)
- log(rel.var.building density + 1)^2
- log(elevation+ 1)
- log(elevation + 1)^2
- rel.var. pop. Density
- (rel.var. pop. density)^2
- log(distance from roads + 100)
- log(distance from roads + 100)^2
+ conversion of land use
+ distance from urban area
+ (distance from urban area)^2
+ log(slope + 1)
+ log(slope + 1)^2

It can be observed that slope and land use conversion have been removed from the original model.

The equation terms selected by the stepwise procedure are entered to fit the GLM by applying formula (8). The following β coefficients are estimated and it is achieved a D^2 of 0.27:

Table 6 – Result from calibration process.
(*) indicates the level of importance of the equation term which is based on the p-value.

	Estimate	Std.Error	z value	Pr(>|z|)	
(Intercept)	50.039487	9.310065	5.375	7.67E-08	***
log(distance from periurban area + 100)	2.399218	2.13322	-1.125	0.26072	**
log(distance from periurban area + 100)^2	-0.220236	0.179405	1.228	0.2196	**
log(rel.var.building density + 1)	5.114541	1.90826	2.68	0.00736	**
log(rel.var.building density + 1)^2	-7.820861	3.044721	-2.569	0.01021	*
log(elevation+ 1)	8.453754	1.540952	5.486	4.11E-08	***
log(elevation + 1)^2	-0.931544	0.176825	-5.268	1.38E-07	***
rel.var. pop. Density	0.107315	0.040716	2.636	0.0084	**
(rel.var. pop. density)^2	-0.003843	0.001339	-2.871	0.00409	**
log(distance from roads + 100)	21.929387	3.151034	-6.959	3.42E-12	***
log(distance from roads + 100)^2	1.887691	0.279367	6.757	1.41E-11	***

Legend:
rel.var.building density : relative variation of building density
rel.var. pop. density: relative variation of population density
log: logarithm
^2: quadratic function

The estimation of unknown β coefficients, represents the goal of regression analysis. Coefficients indicate how the change in one of the independent variables affects the values taken by the dependent variable. To return the expected value (Y), i.e. the probability of building occurrences, the linear predictors must be transformed into the scale of the response

value. This is performed by applying the inverse of the Logit link function for binomial distribution (table 3); operation that is implemented in ArcMap. The following two images show the output map resulting from the calibration of the model. The first one (fig. 27) displays the probability of building occurrences across the entire rural area. In fact, even if the model is calibrated on a dataset acquired from sample areas, since these are representative portion of the study area, the probability estimation of building occurrences embraces the entire rural area. The second images shows the overlay between the map of probability elaborated by the model and the actual rural built environment existent in 2005 (fig. 28).
Probability map of building occurrence is a grid where to each cell is associated the value of probability resulting from the running of the calibrated model.
GIS tools allow to display the resulting map and to evaluate visually the goodness of fit of the model. However for a better accuracy and assessment, the fit between probability of building occurrences, estimated by the model, and buildings actually existing in 2005 is evaluated by performing the mean of probability values of cells whose location is inherent to actual building features. In fact, it is expected that cell values located in correspondence of existing buildings, if properly estimated, have probability value equal to 1. Conversely if the cell value is not correctly estimated, has low values of probability. The measure of the mean performed on these collected cells values, provides a useful assessment tool of the goodness of the model. A low mean score must be interpreted as an incapability of the model in estimating building allocation.

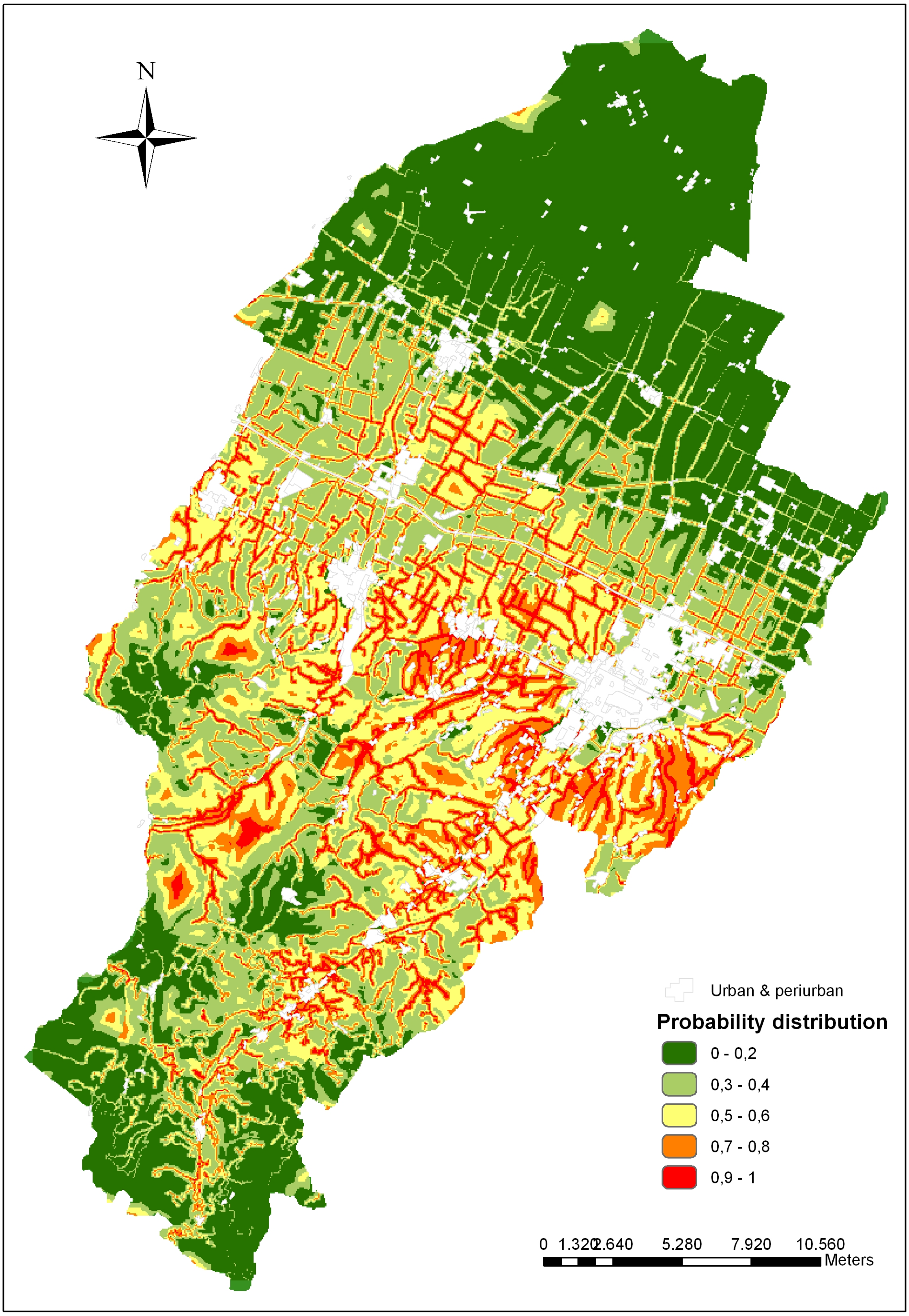


Figure 27 - Probability distribution map of building occurrences inside the rural area.

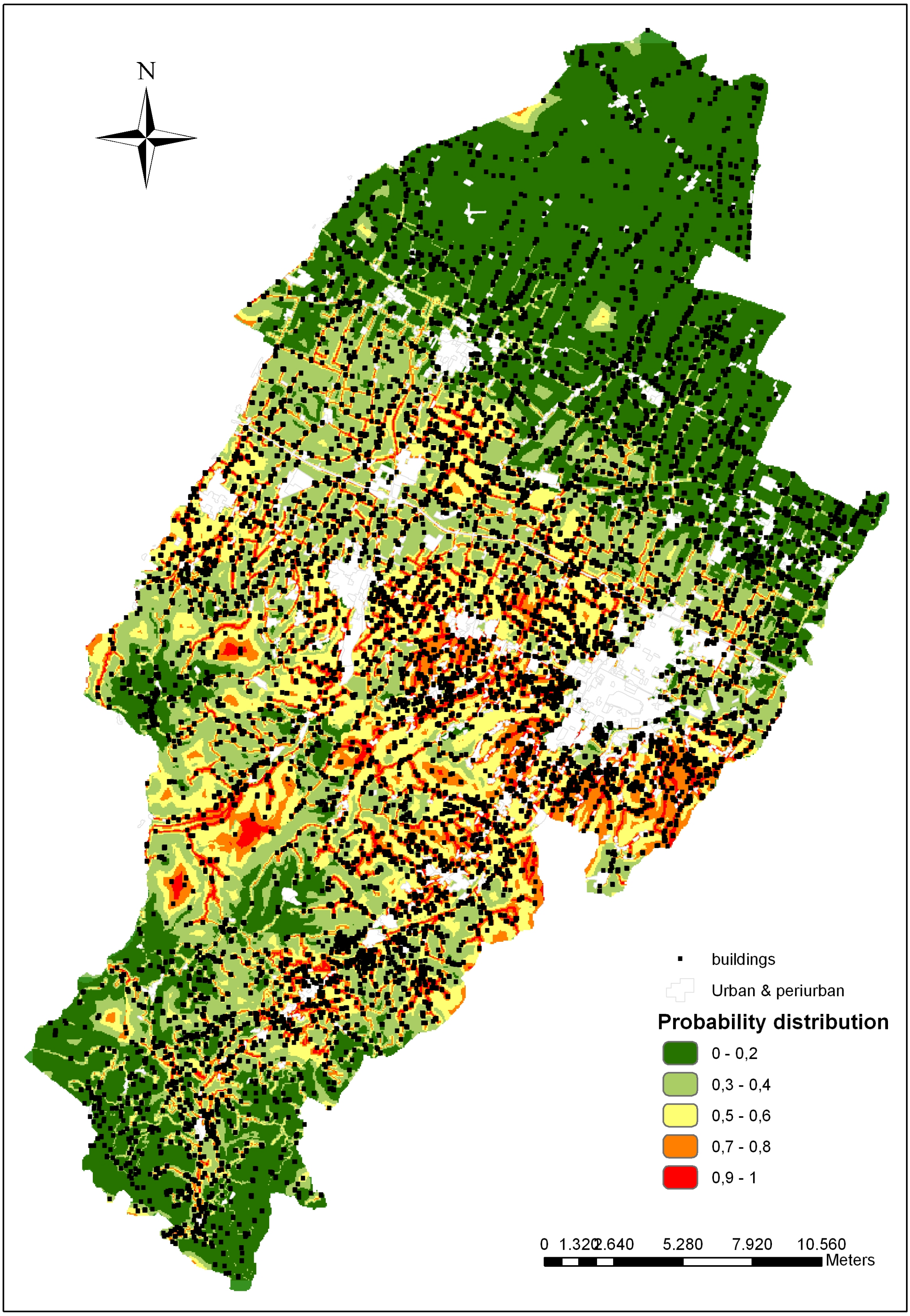


Figure 28 - Probability distribution map with the existing buildings in 2005 inside the rural area.

Once again with the Hawth's tool is performed the spatial overlay between layer of buildings in 2005 within the rural area and raster probability map and then the intersection between points representing building allocations in 2005 and the corresponding cell values of probability of building occurrence. As a result a new field is attached to the attribute table of "existing buildings in 2005" displaying for each point feature the inherent value ranged from 0 to 1 of probability values estimated by the model.

This process was carried out separately for each town of the New District of Imola, so to provide a more detailed frame of driving forces considered in the model.

For proper work of the method it is not sufficient to verify only the spatial correspondence between pixels with high probability and the built environment actually existent, but also between low probability value and the absence of the built environment. In this way it can be verified not only if the model is able to correctly predict locations where actual built environment exists, but also those locations where is not existent (fig.29).

Table 7 – Mean of probability values per town.

	PRESENCE		ABSENCE	
	Mean	**St.dev**	**Mean**	**St.dev**
MORDANO	0.7	0.2	0.1	0.1
IMOLA	0.5	0.3	0.3	0.2
DOZZA	0.7	0.2	0.6	0.2
CASALFIUMANESE	0.7	0.3	0.4	0.2
FONTANELICE	0.7	0.2	0.3	0.2
CASTEL DEL RIO	0.4	0.2	0.1	0.2
CASTEL SAN PIETRO TERME	0.7	0.2	0.4	0.2
CASTELGUELFO	0.7	0.2	0.4	0.2
BORGO TOSSIGNANO	0.7	0.2	0.4	0.1
MEDICINA	0.2	0.2	0.1	0.1

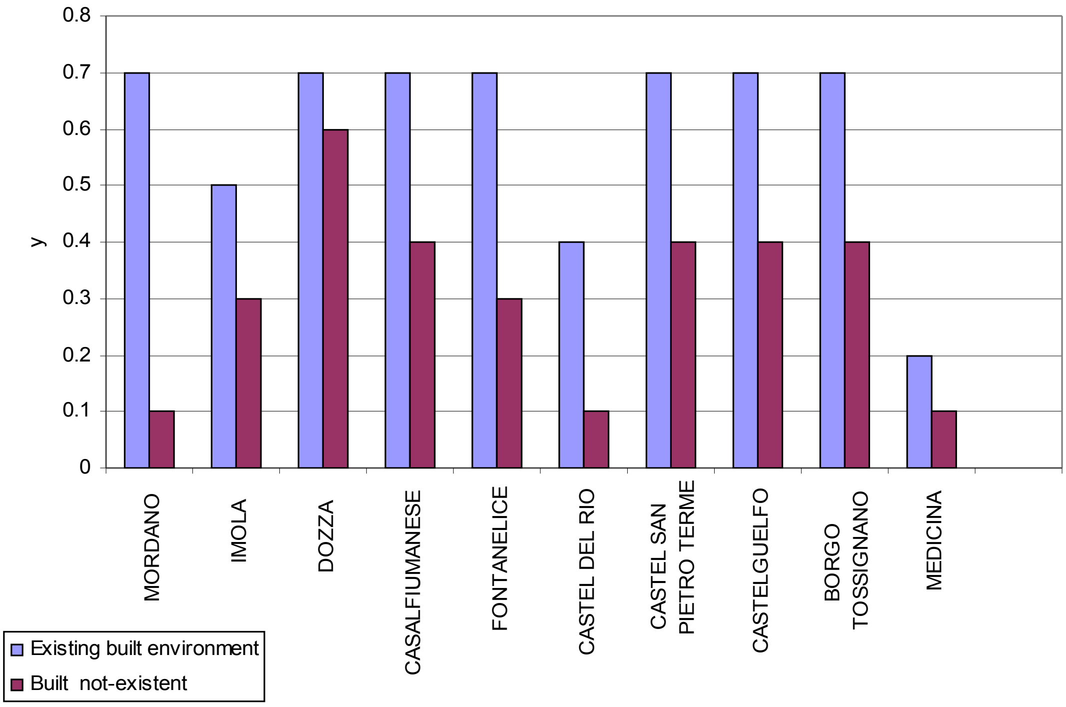


Figure 29 - Means of probability cell values in correspondence of actual built environment and not existent built environment.

Observing the mean scores it can be noticed that the probabilities estimated by the model converge in the most part of municipalities managing to capture the presences of building actually existent and absences. The only towns where model shows a lack of reliability in its power of estimating the occurrence of buildings are the towns of Medicina and Castel del Rio where mean scores are close to 0.2 and 0.4. The prediction power in these municipalities is low because the model can not record presences of the built environment. This behaviour can be explained observing the distribution of variables values along the study area (see fig 30; 31; 32). It can be notice that some variables entered in the model, which introduce a high degree of variability to the model, led to a low probability estimation in these areas. In particular the distance from periurban area is effective in recognizing the possibility of existence of the built environment in areas located northern of Medicina (see fig.30). It might be other driving forces able to explain the presence of these settlements, which were not expressed by the variables included in the model.

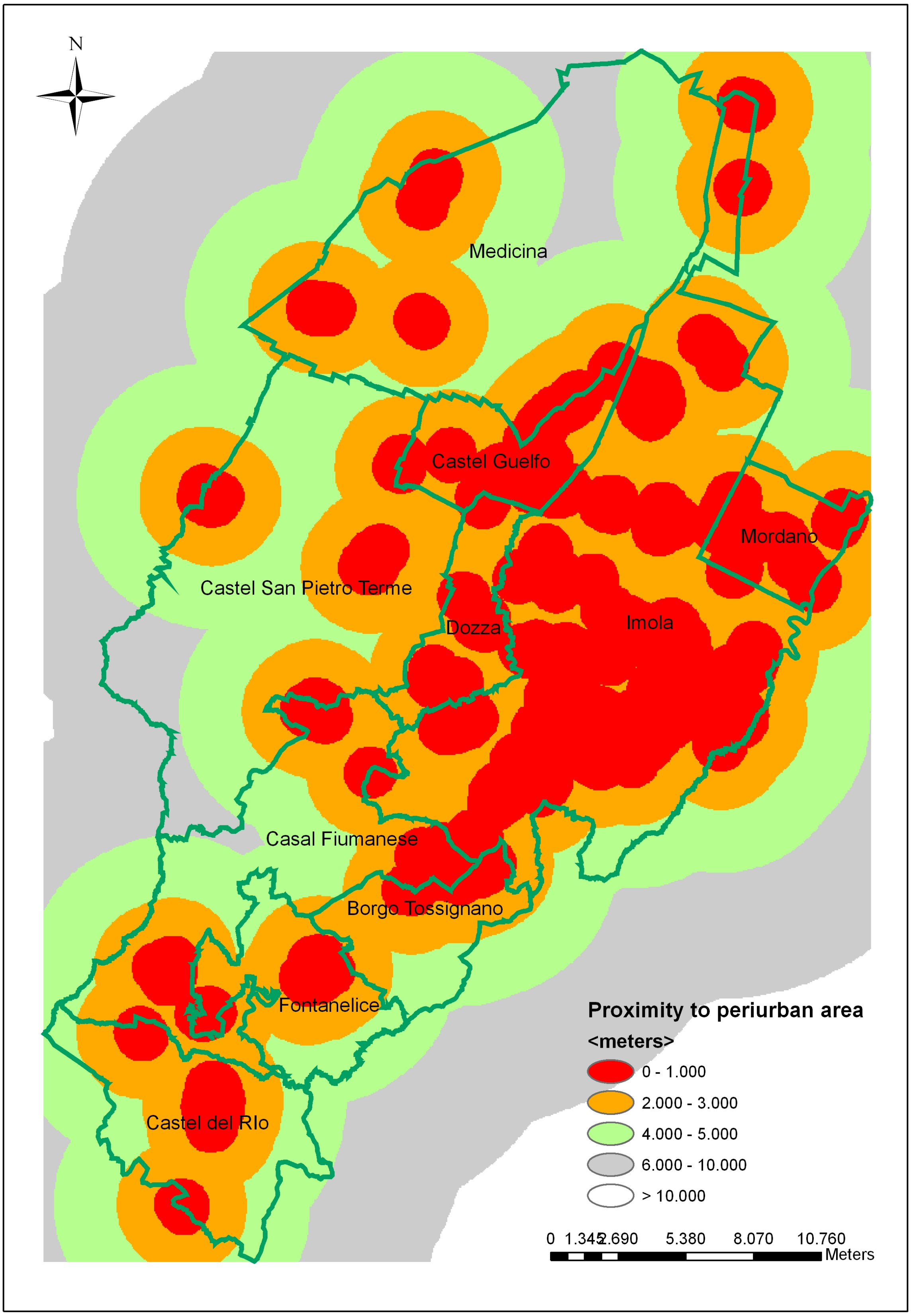


Figure 30 – Value distribution of explanatory variable: distance from periurban area.

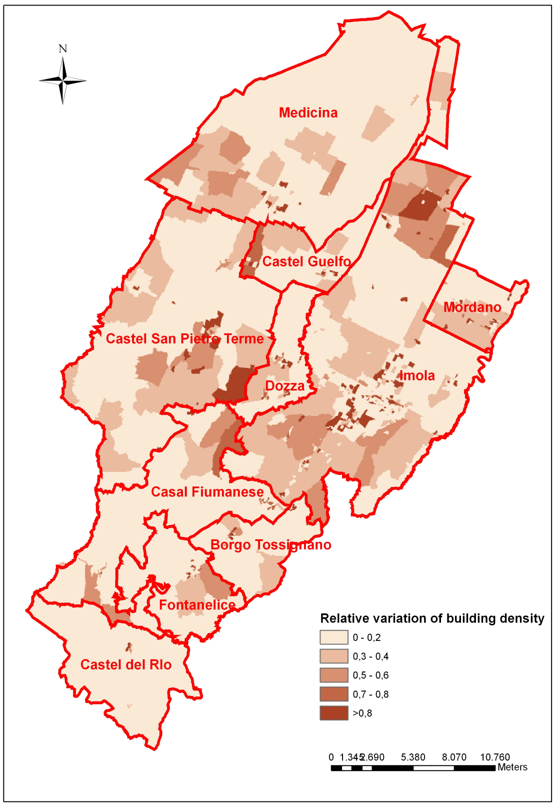


Figure 31
Relative variation of building density

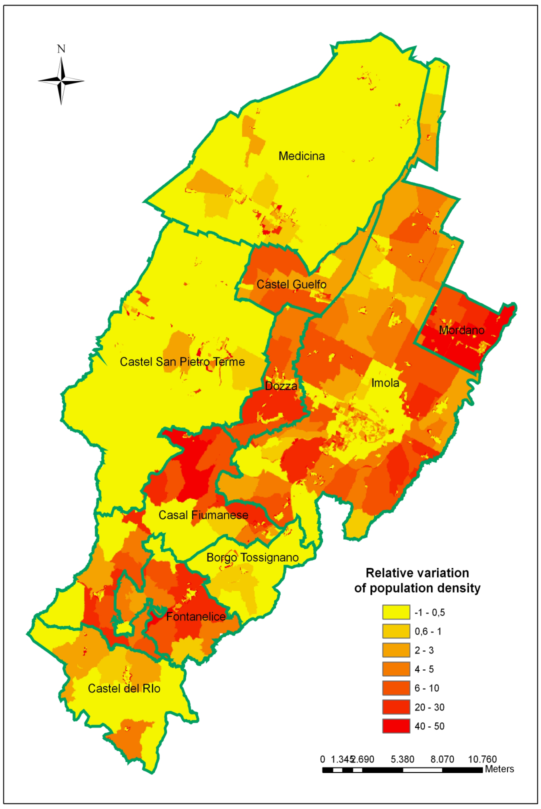


Figure 32
Relative variation of population density

Some interpretations can be attempted on explanatory variables to evaluate their influence on the building expansion process. Since explanatory variable after adjustment transformation are expressed in the model by equation terms, it is necessary in first place solve such equations in order to evaluate their behaviour in the model (see fig. 33).

It can be noticed that the expansion of building decreases further away from the periurban area and the same dynamic occurs moving away from roads. There is an increasing of expansion of buildings close to values of relative variation of population density that once reached a saturation values tends to decrease. The elevation variable expresses how building expansion benefits from flat terrain, and starting from value close to 200 m there is a decrease of built-up area.

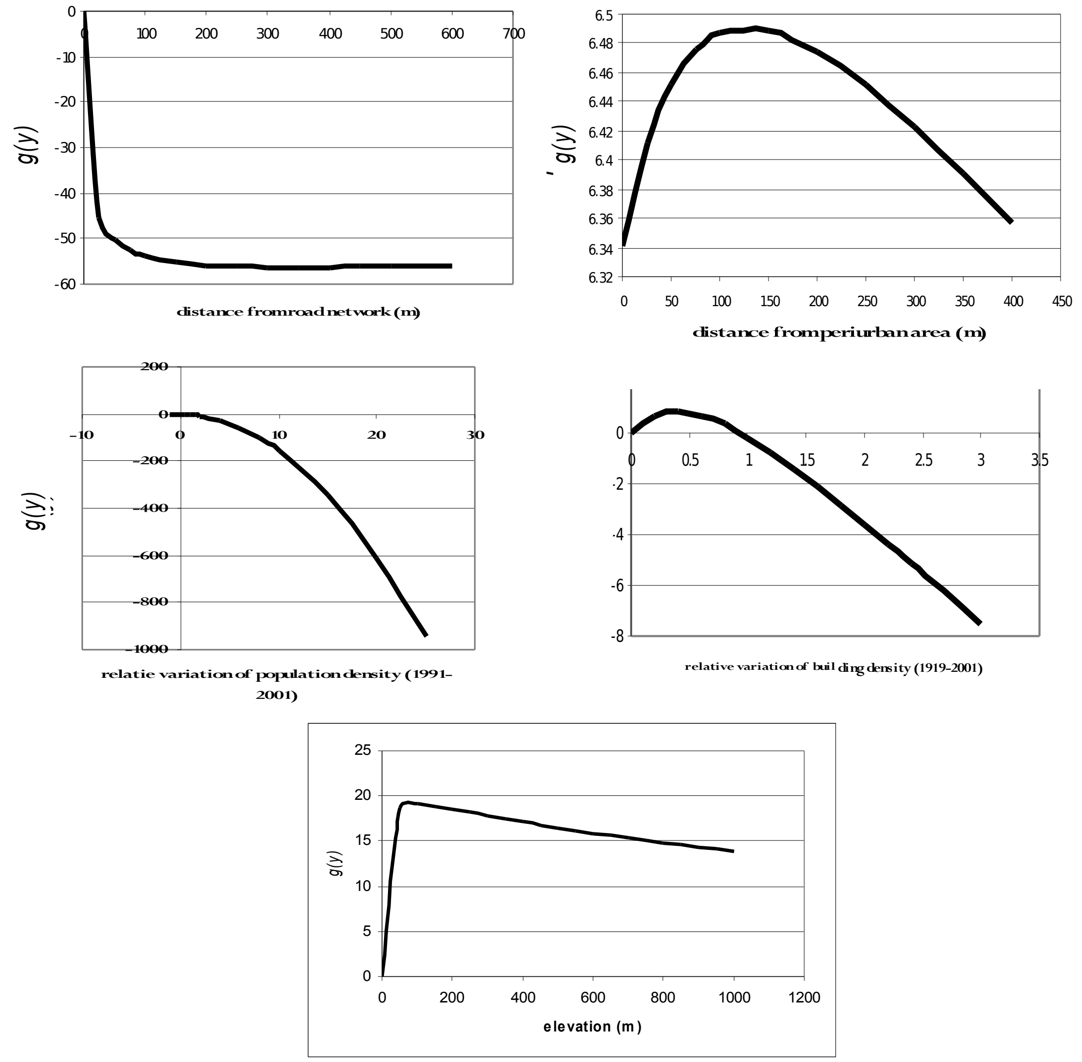


Figure 33 – Trends of explanatory variables express in the scale of linear predictors

After the calibration process it becomes aware how much the model explains of the initial overall deviance. Although it’s not declared how much each predictor contributes individually to the model. Until recently, basically no tool was available to assess these independent contributions. Anova tables were currently used, but they cannot be interpreted correctly unless the predictors are fully independent of each others.

One way to better estimate the relative contribution of each predictor in a GLM model is thus to fit models for all possible combinations of predictors, and to calculate the “mean contribution” of each predictor across all models. This complex computation can be done in R with the command *hier.part.*

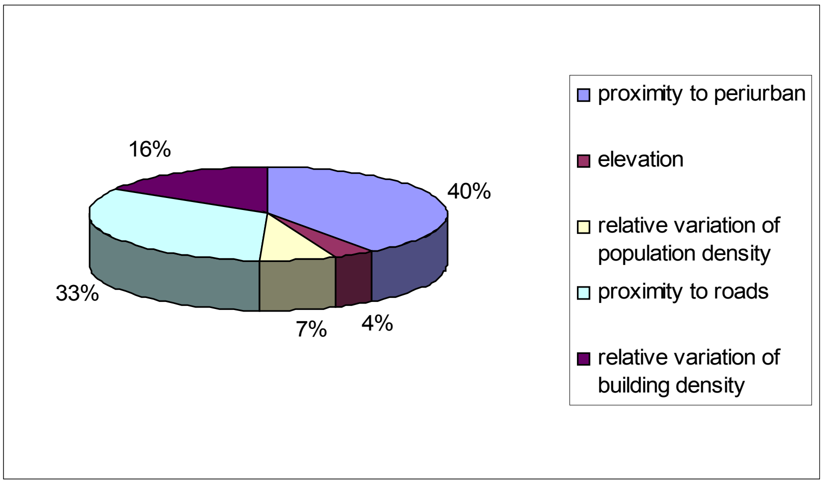


Figure 34 - Independent effects of explanatory variables expressed in percentage.

Independent values express the ratio of variability imported by each variable on the overall variability explained by the model, and the distance from periurban area together with the distance from road network as shown in fig. 34 provide the most significant contribution to the variability of the model.

Another test that was performed on the model, this time to verify the accuracy of the model in fitting the observed data by means of the cross-validation test. As explained in paragraph III.2.8 The adoption of such test is due to the restricted number of observations. Hence the evaluation test provides a methodology that allows to test the model on the base of the same dataset used for the calibration process.

The evaluation metric adopted to express the level of accuracy is the AUC, the area under the curve of a ROC (Receiver Operating Characteristic)-plot, for explanations regarding the description of the method it is remained to paragraph III.2.8.

The AUC obtained is 0.75 which expresses a useful prediction of the model.

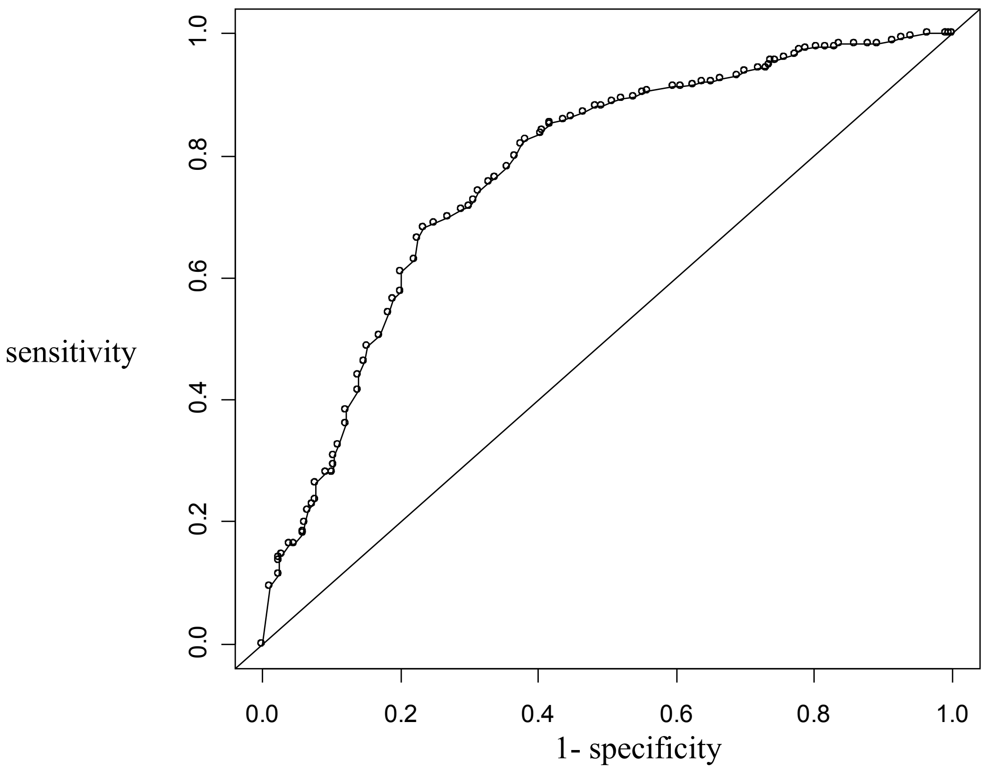


Figure 35 - ROC-plot from model of rural area with AUC = 0.75.

IV.2. MODEL CALIBRATION OF PERIURBAN AREA

A different model is built on the base of the periurban data set, to explain building allocation within periurban area. Such approach is necessary, as explained in the chapter related to material and methods (III.2.5), because driving forces acting on periurban building expansion are considered to be different to those acting on rural area.

It follows the list of explanatory variables recognized as providing significant contribution to the building allocation process. Their selections are motivated in paragraph III.2.5.

1. Distance from roads
2. Relative variation of building density
3. Relative variation of population density
4. Distance from urban area
5. Slope
6. Elevation

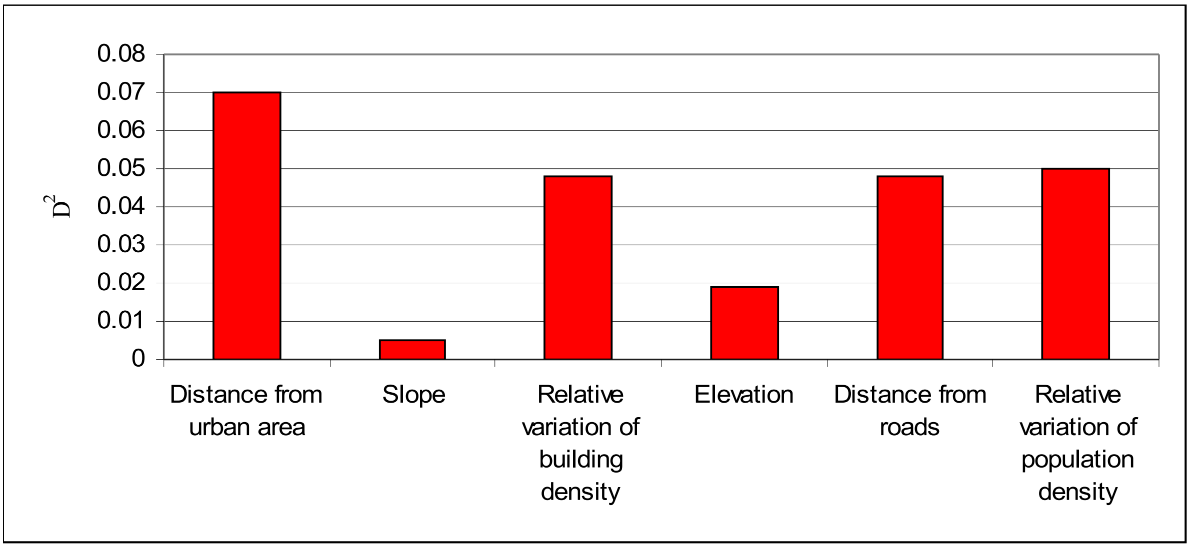


Figure 36 - D^2 values of explanatory variables before the adjustment process.

Also in this case, to fit the periurban model some adjustments of equation terms are attempted to achieve better D^2 values.

Before adjustment:

$D^2 = 0.13$

After adjustment:

$D^2 = 0.22$

Table 8 – Explanatory variables that show some significant improvements of D^2 values obtained with the adjustment process.

	D^2 before adjustment	D^2 after adjustment
Relative variation of building density	0.041677735 →	0.0650256
Distance from urban	0.071067636 →	0.0739861
Distance from road network	0.048493420 →	0.1200099

Mostly log or quadratic transformations are performed. Terms with the best D^2 are entered into the GLM model applying formula (8). The output from the calibration provides β coefficients as shown in the table below.

Table 9 – Output from the first calibration process.

	Estimate	Std. Error	z value	Pr(>|z|)	
(Intercept)	-3.18E+00	8.56E-01	-3.711	0.000206	***
log(distance from roads + 1)	1.85E+00	2.72E-01	6.793	1.10E-11	***
(log(distance from roads + 1)^2)	-2.84E-01	4.12E-02	-6.888	5.68E-12	***
log(vr_bls_dns + 0.001)	2.62E-01	5.61E-02	4.672	2.98E-06	***
(log(vr_bls_dns + 0.001)^2)	2.46E-02	9.19E-03	2.679	0.007391	**
log(distance from urban + 1)	8.98E-01	3.25E-01	2.759	0.005795	**
(log(distance from urban + 1)^2)	-1.52E-01	4.25E-02	-3.59	0.000331	***
(slope)^2	9.73E-02	8.71E-02	1.117	0.263981	
(vr_pop_dns)^2	-1.80E-01	1.13E-01	-1.592	0.111342	**
(slope)	-1.48E-02	3.44E-02	0.431	0.666661	
elevation	8.73E-03	8.83E-03	0.988	0.323316	
I(elevation^2)	-3.38E-05	3.32E-05	-1.018	0.308902	

The selection of the most significant explanatory variables among those entered is performed by using the stepwise procedure. The selection process provides the following output:

Table 10 - Output from stepwise selection:
(-) selected variables; (+) discharged variables.

- log(distance from roads + 1)
- (log(distance from roads + 1)^2)
- log(vr_bls_dns + 0.001)
- (log(vr_bls_dns + 0.001)^2)
- log(distance from urban + 1)
- (log(distance from urban + 1)^2)
- slope
- vr_pop_dns
+ (slope)^2
+ elevation
+ (elevation^2)

It can be notice that both terms that own elevation variables have been rejected from the final model. This means the irrelevant contribution of such variable to the variability of model.
From the second calibration process, run with equation terms selected by the stepwise procedure*,* new coefficients were generated as shown in the table below:

Table 11- Output from calibration process after stepwise selection.

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	-2.83E+00	7.82E-01	-3.611	0.00031 ***
log(proximity to roads + 1)	1.85E+00	2.70E-01	6.843	7.77E-12 ***
(log(proximity to roads + 1)^2)	-2.83E-01	4.07E-02	-6.968	3.21E-12 ***
log(vr_bls_dns + 0.001)	2.60E-01	5.52E-02	4.71	2.47E-06 ***
(log(vr_bls_dns + 0.001)^2)	2.46E-02	9.12E-03	2.698	0.00698 **
log(proximity to urban + 1)	9.15E-01	3.24E-01	2.822	0.00477 **
(log(proximity to urban + 1)^2)	-1.55E-01	4.22E-02	-3.669	0.00024 ***
(slope)	-1.43E-01	5.44E-02	2.634	0.00845 **
(vr_pop_dns)^2	-1.84E-01	1.13E-01	-1.633	0.10247 **

The map of probability distribution is obtained by applying the inverse of the link function, in fact as previously explained, the response variable in the GLM is computed in the same scale of the linear predictors (i.e. explanatory variables). The map of building probability distribution for the periurban area is obtained by solving equations reported in table 3 in paragraph III.2.7 of binomial distribution. As for the rural map of probability of building occurrence, also for the periurban map, the output projects the probability of occurrence on the entire surface of the study area, because explanatory variable values are provided on the whole territory, hence the calculation is performed for cell pixel value of the study area (fig 37, 38).

However, only for the periurban area the probability map of building occurrence is representative and hence the output map is clipped along periurban perimeter as shown in the image below.

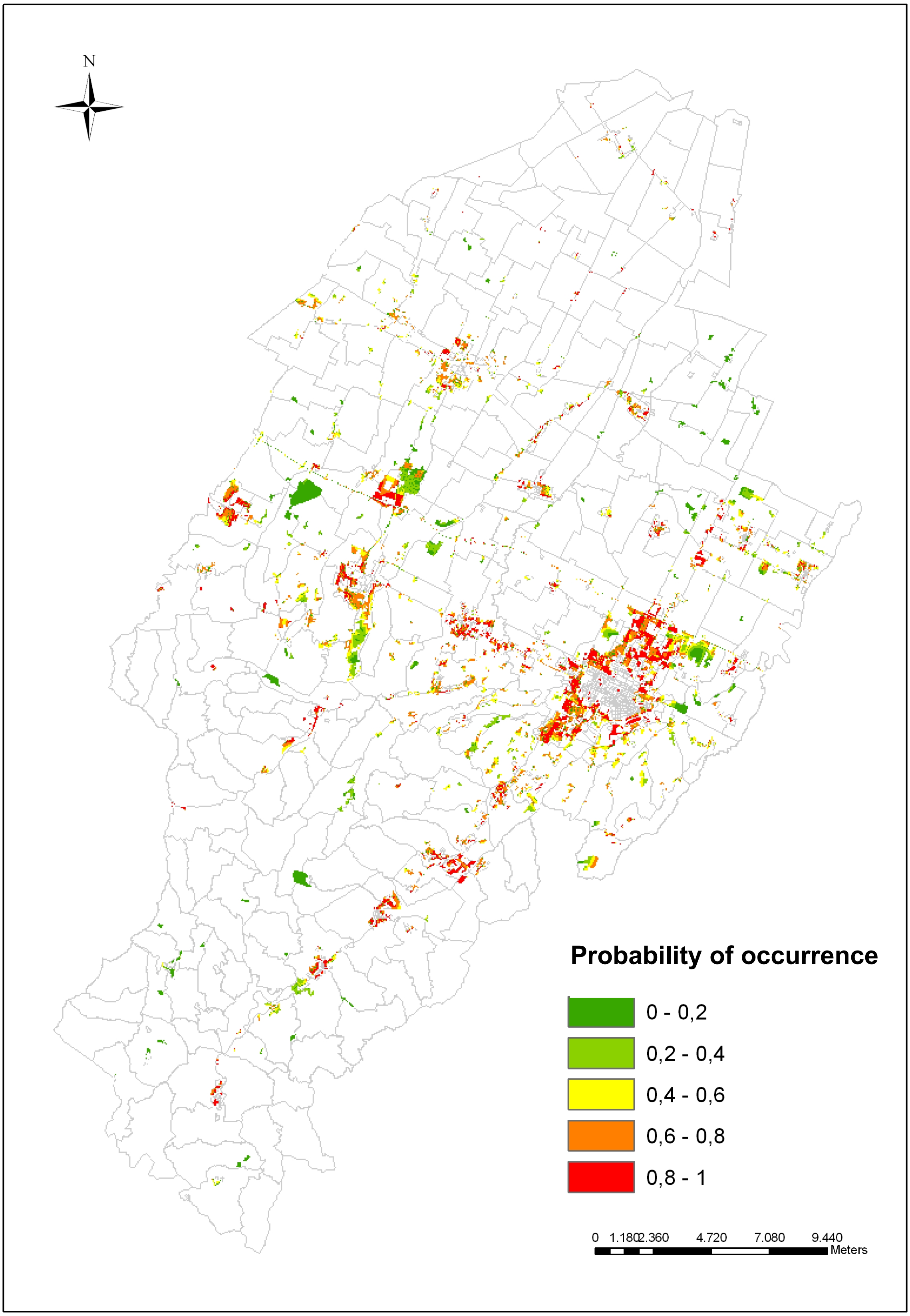


Figure 37- Map of building probability distribution in periurban area.

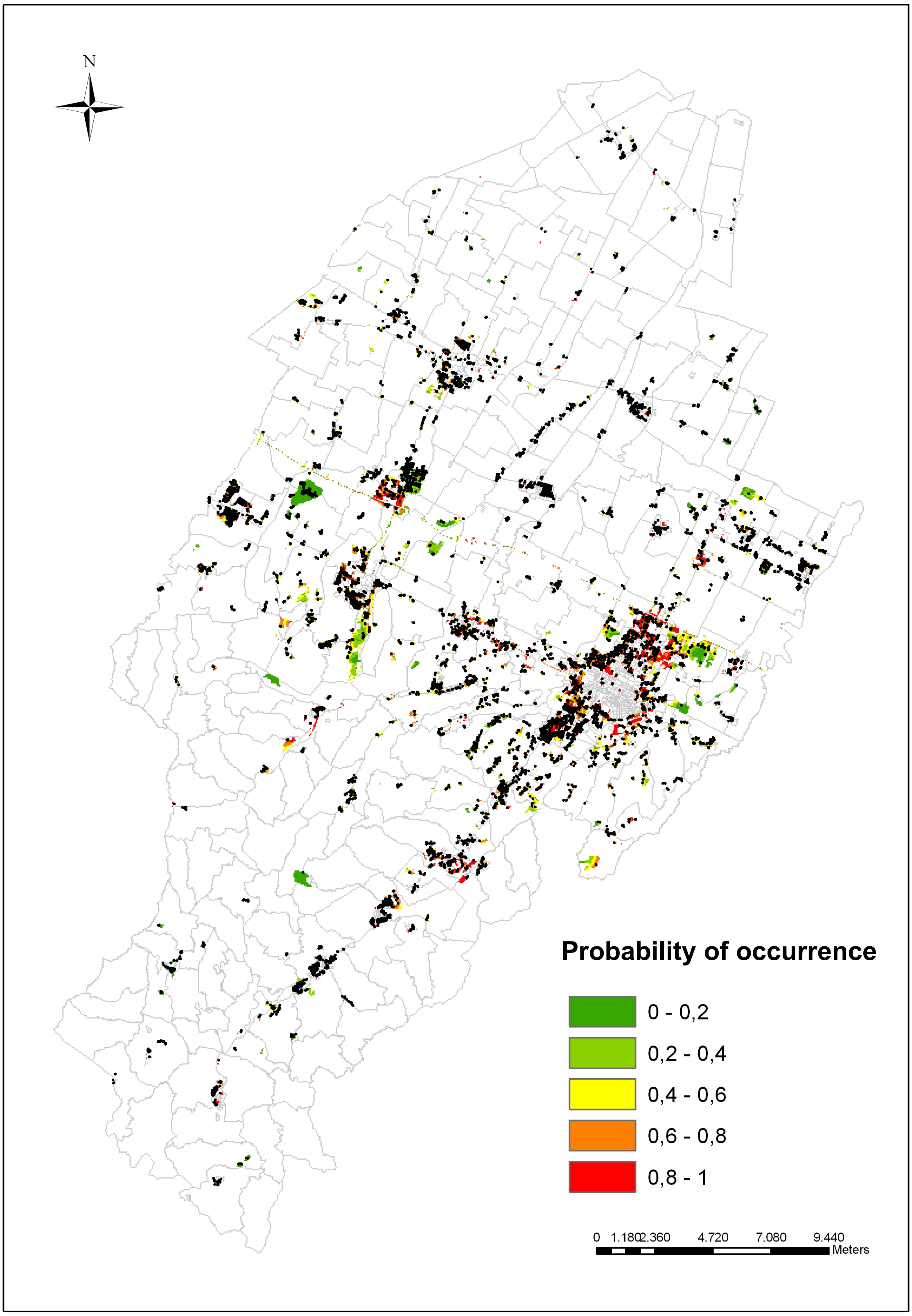


Figure 38 - Map of building probability distribution in periurban area with existing built environment in 2005.

The output map (fig.39,40) displays high probability occurrence in the closest surrounding of the greater urban area, hence the distance from urban must have been a strong predictor power.

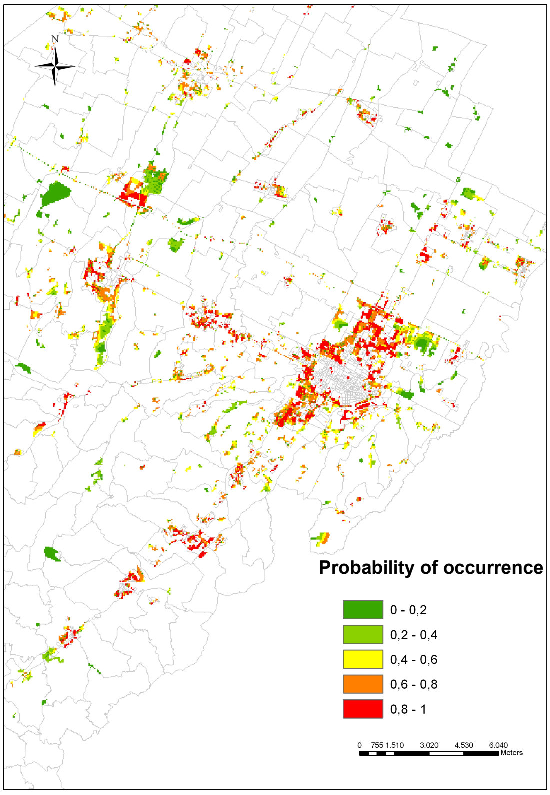


Figure 39 - Probably distribution map of buildings in periurban area: without rural buildings.

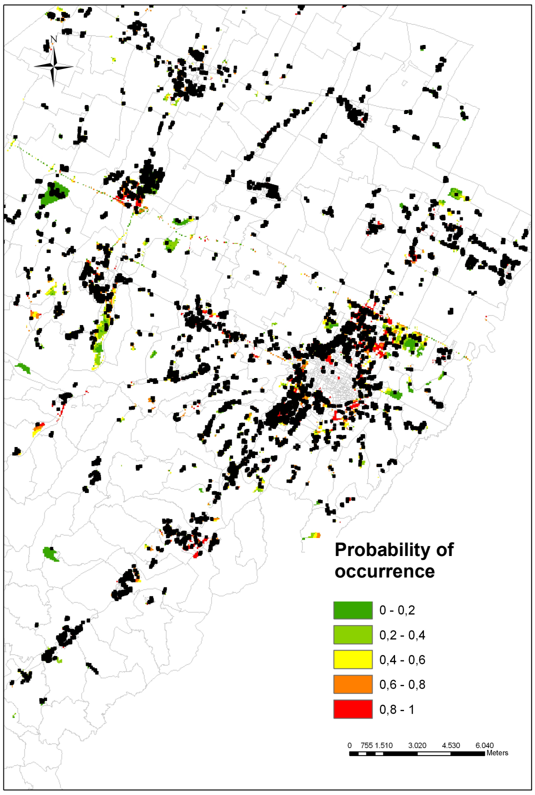


Figure 40 - Probably distribution map of buildings in periurban area: with existing rural buildings in 2005.

Also for the periurban model the same test of the means is applied to asses the estimation capability of model.

Table 12 – Mean and Standard deviation of probability values per town of periurban model.

	PRESENCE		ABSENCE	
	Mean	**St.dev**	**Mean**	**St.dev**
BORGO TOSSIGNANO	0.7	0.2	0.4	0.3
CASALFIUMANESE	0.3	0.2	0.1	0.2
CASTEL DEL RIO	0.7	0.2	0.3	0.3
CASTEL GUELFO	0.7	0.2	0.2	0.1
CASTEL SAN PIETRO TERME	0.7	0.2	0.3	0.3
DOZZA	0.7	0.2	0.3	0.3
FONTANELICE	0.6	0.2	0.2	0.2
IMOLA	0.8	0.2	0.3	0.3
MEDICINA	0.6	0.2	0.3	0.3
MORDANO	0.7	0.2	0.2	0.2

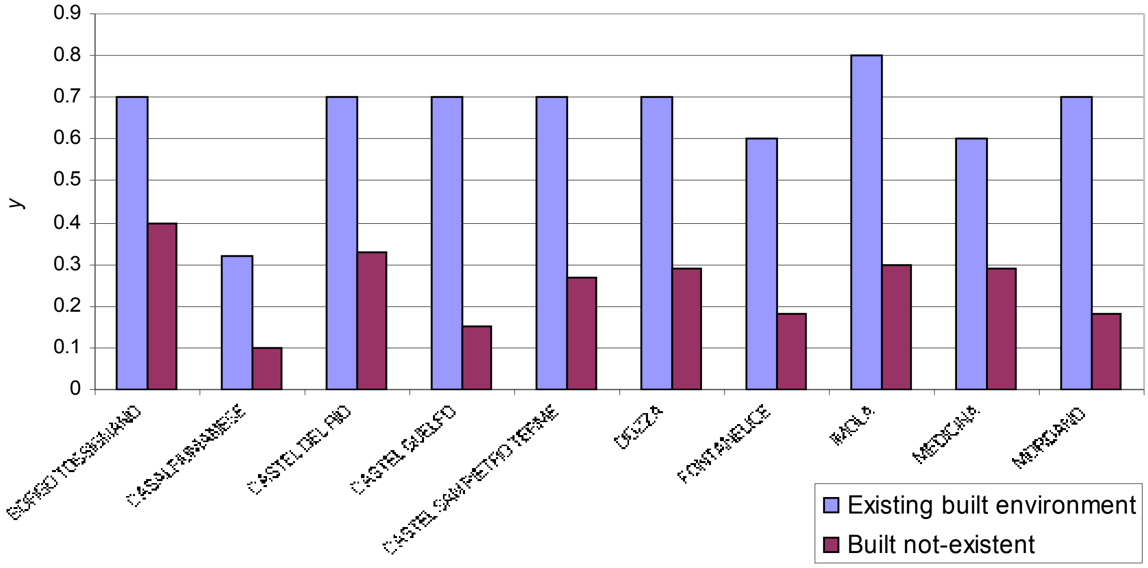


Figure 41 - Mean values of probability pixel values in correspondence of actual built environment and not existent built environment.

It can be noticed that means performed on probability values associated to the existing built environment per town show quite high score. Hence it can be said that the model managed to capture the most part of presences of building actually existent. The only town where model demonstrated a lack of reliability in its power of estimating building occurrences is the town of Casalfiumanese, which doesn't own a large urban centre. However, observing the distribution map of relative variation of population density values, it can be noticed how in correspondence of the periurban area of Casalfiumanese there are very low cell values (see fig.42, 43).

The absence of built environment is captured by model and this is shown in the table 12 by examining the low values of the mean per town.

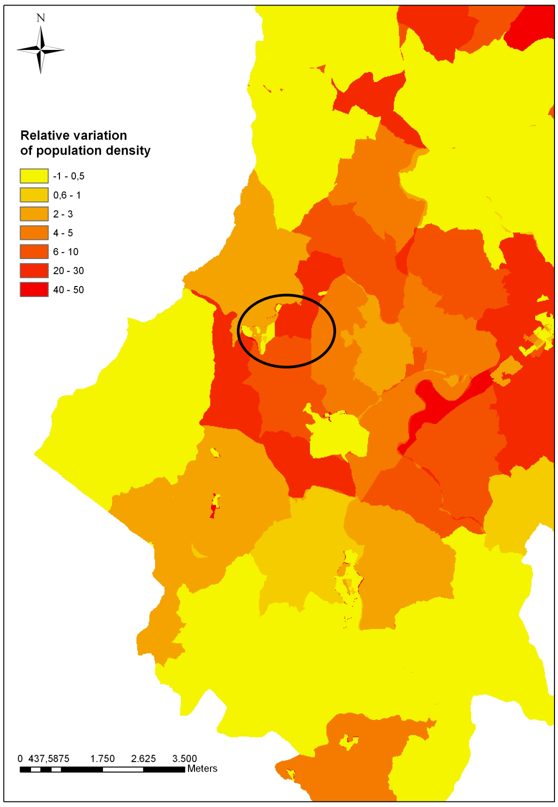


Figure 42 – Relative variation of population density

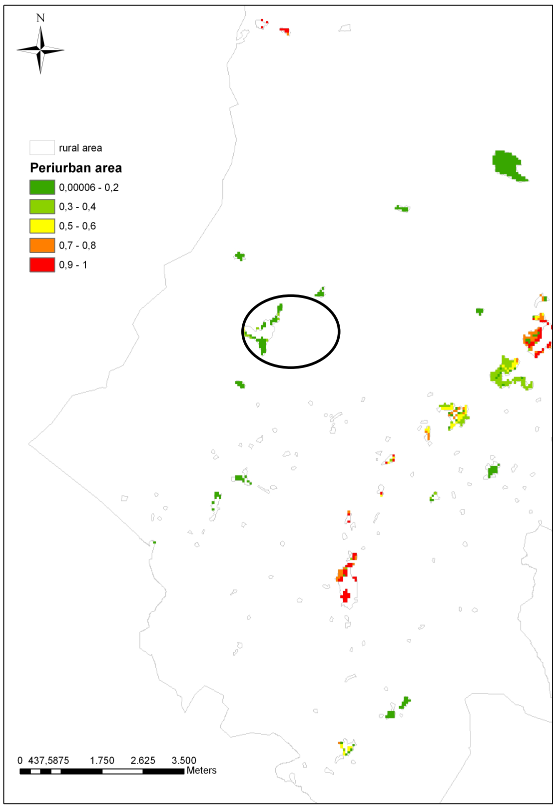


Figure 43 – probability of building occurrence in periurban area

Some interpretations can be attempted on explanatory variables to evaluate their influence on the building expansion process. Since explanatory variables after adjustment transformation are expressed in the model by equation terms, it is necessary in first place to solve such equations in order to evaluate their behaviour in the model (see fig. 44).

The allocation of buildings decreases proceeding further away from the urban area and the same dynamic occurs moving away from roads. It is shown an increasing of expansion of buildings at the increasing of relative variation of population density and of relative variation of building density. The elevation variable expresses how building expansion benefits of flat terrain and for altitude values close to 200 there is a decrease of built-up area.

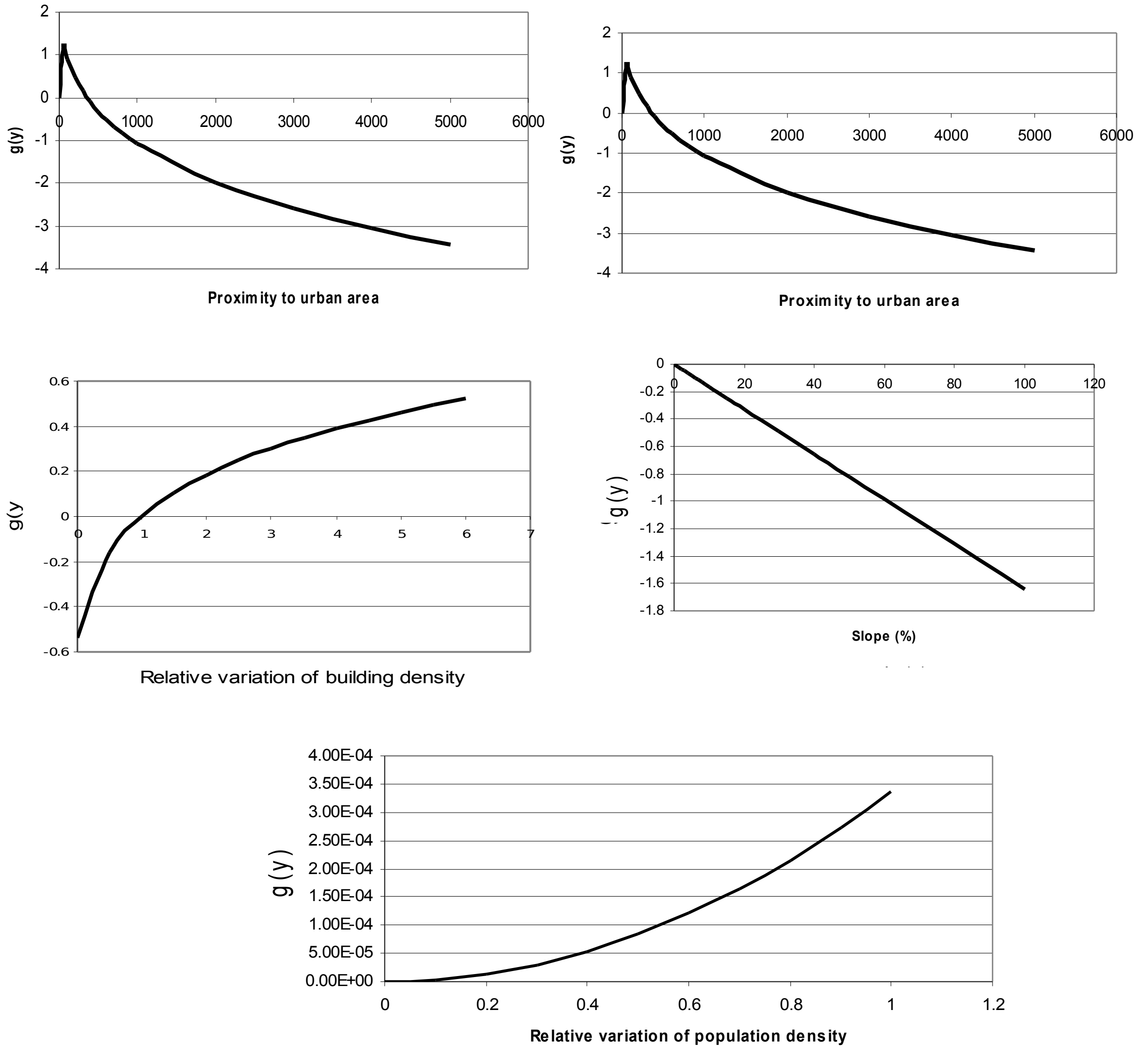


Figure 44- Trends of explanatory variables expressed in the scale of linear predictors.

Beta coefficients represent the quantitative contributions of each independent variable into the prediction of the dependent variable. However to evaluate how much each predictor contributes individually to the GLM model it is necessary to fit models for all possible combinations of predictors, and to calculate the “mean contribution” of each predictor across all models (fig.45). The same was performed for model of rural area with the command *hier.part*.

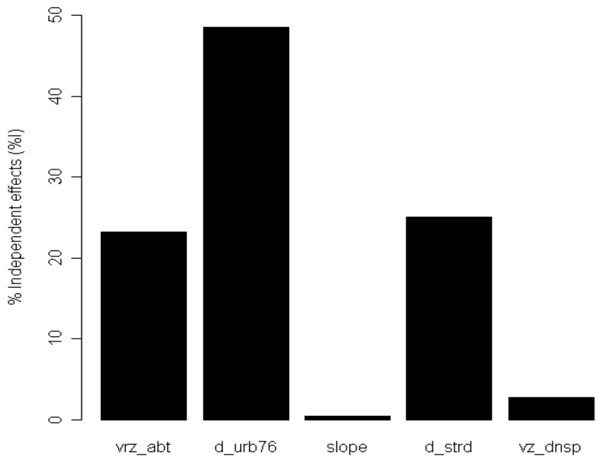


Figure 45 – Indipendent effects of explanatory variables

The goodness of model is then tested with the cross-validation, since as for the rural dataset, number of observations is not sufficient to leave some data out for the test process. The AUC metrics resulting from the test is 0.77 which expresses the useful prediction of the model (fig.46).

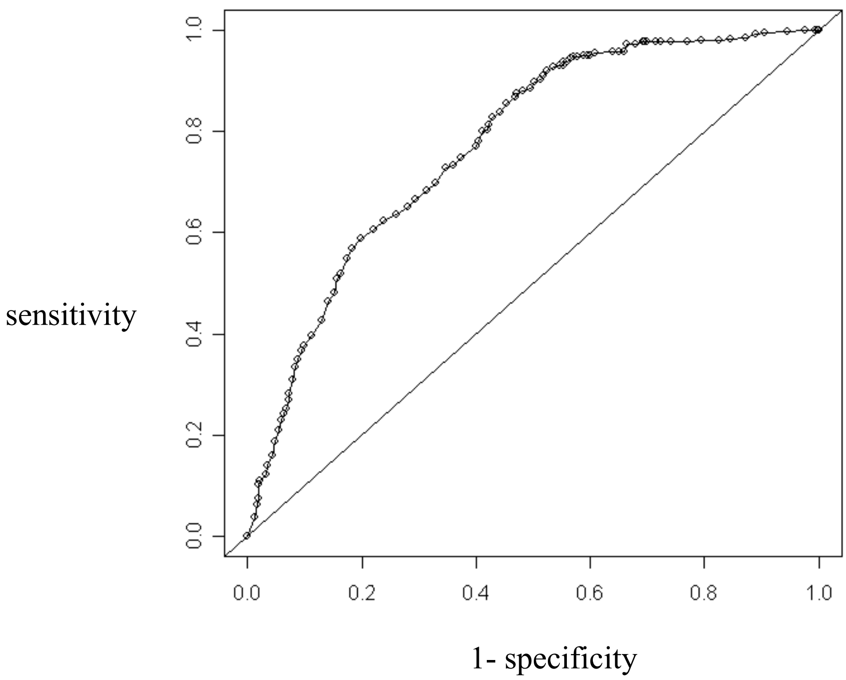


Figure 46- ROC-plot of periurban model.

V CONCLUSIONS

V. CONCLUSIONS

A careful review of scientific literature, with regard to rural landscape scenario, reports a wide range of models aimed at the comprehension of driving forces acting in land-use change dynamics, landscape transformations, and urban expansions but underlines a general lack of models for the understanding of driving forces acting on rural built environment.
The study of driving forces is believed to provide useful tools in planning policies for the decision making process, since they allow to have insight in past territorial dynamics that can support the formulation of possible future scenarios. In particular, since some of the main issues affecting rural landscape are related to an irrational and intensive use of lands for the allocation of new structures and infrastructures related to human activities, the methodology developed focused on the definition of a spatial explicit model able to asses causal-effect relationships between possible driving forces and the allocation process of new buildings.

The methodology proposes a set of steps that lead to the definition of a spatial model. On this purpose required statistical techniques currently available to carry out elaborations were analyzed. In particular the methodology seeks to developed a spatial model based on the most suitable statistical technique implemented with GIS software. The study proved the fundamental role of GIS techniques in the development of the spatial model. They demonstrated to be relevant not only in mapping explanatory variables by means of spatial topological relations among spatial features but also in processing the final model.
Particular attention was paid on the testing of the model resulting from the proposed methodology. The spatial model was tested in a case study of building allocation analysis within the New District of Imola. Logistic regression was used to calibrate and interpret the model and the model is then compared and evaluated on the base of the existing building allocation.
Results show that the model has a sound capability in detecting the existence of hypothesised relationships between the driving forces and new building allocations. In particular for some municipalities of the study area the model shows a better prediction power.
The study demonstrates that: (i) the model can be developed based on hypothesised relationships based on consideration of underlying and proximate causes of change, (ii) the model developed by the proposed methodology is able to detect main driving forces and understand the type of relationship with the building expansion providing valuable insights in underlying causes of change occurred within the study area (iii) the model allows to detect not

only the existence of causal relationship between the building allocation process and possible driving forces, but also the type of correlation and the contribution of each variable on the overall variability detected.

Despite the test of the methodology required the assumption of a study area, the proposed method is flexible and applicable to different contexts. In fact, the model can be also applied to the interpretation of the trends which occurred in other study areas, and also referring to different time intervals, depending on the availability of data.
The use of suitable data in terms of time, information, and spatial resolution and the costs related to data acquisition, pre-processing, and survey are among the most critical aspects of model implementation. In fact, some of the most complicated stages in the developing of the model are represented by the parameterization and measurement process of driving forces. Difficulties are due to the impossibility of finding appropriate parameters that allow the expression of driving forces, particularly those concerning cultural or political aspects.
Also the scarcity of data referring to past conditions represents another limit to the possibility of introducing driving forces.
The availability of further data able to express transformation dynamics and hence the possibility to enter others significant driving forces can improve rate of variability explained through the model.
The acquisition of spatial information related to building locations in past decades implied the application of a random stratified sampling methodology. The sampling methodology itself, is part of a wider research study, and the spatial information regarding buildings, that were used as observation for the model, represent the result of the application of such methodology to the development of a case study.

The modest amount of observations on spatial building allocation forced the application of Jack-knife as validation test. Moreover the fact that the model was developed on the base of observations acquired from extracted sample areas, required the setting of a technique for the goodness assessment of the model fit.

One of the most relevant application of the spatial model is represented indeed by the development of future scenarios.
Maintaining invariant estimated coefficients, obtained from the model calibration, and updating values of explanatory variables it is possible to generate from the running of the

model, a probability map displaying the estimated probability of building occurrences within the same study area in the future.

Future in-depth studies can focus on using the proposed model to predict short/medium-range future scenarios for the rural built environment distribution in the study area. In order to predict future scenarios it is necessary to assume that the driving forces do not change and that their levels of influence within the model are not far from those assessed for the time interval used for the calibration.

.

REFERENCES

Andersen, L.E., 1996. The causes of deforestation in the Brazilian Amazon. J. Environ. Dev. 5, 309–328.

Antrop, M. 1997. The concept of traditional landscapes as a base for landscape evaluation and planning. The example of Flanders Region. Landscape and Urban Planning 38, 105-l 17.

Antrop, M. 2000. Background concepts for integrated landscape analysis. Agriculture, Ecosystems and Environment. 77, 17–28.

Antrop, M. 2001. The language of landscape ecologists and planners: A comparative content analysis of concepts used in landscape ecology. Landscape and Urban Planning, 55, 163-173

Antrop, M. 2004. Landscape change and the urbanization process in Europe. Landscape and Urban Planning, 67, 9-26.

Antrop,M., Eetvelde, V.V. 2000. Holistic aspects of suburban landscapes: visual image interpretation and landscape metrics. Landscape and Urban Planning. 50, 43-58.

Aspinall, R.J. 1994. Exploratory spatial data analysis in GIS: generating geographical hypotheses from spatial data. In: Worboys, M.F. (Ed.), Innovations in GIS 1, Taylor & Francis, London. 119–131

Aspinall, R.J. 2004. Modelling land use change with generalized linear models,a multi-model analysis of change between 1860 and 2000 in Gallatin Valley, Montana. Journal of Environmental Management. 72, 91–103.

Austin, M. P. 2002. Spatial prediction of species distribution: an interface between ecological theory and statistical modelling. Ecological modelling. 157, 101-118.

Baker, W.L. 1989. A review of models of landscape change. Landscape Ecology vol. 2 no. 2 pp 111-133 SPB Academic Publishing, The Hague.

Batty, M., Densham, P. 1996. Decision support, GIS and urban planning. ST. Sistema terra 5, 72-76.

Batty, M., Longley P. 1989. Urban growth and form: Scaling, fractal geometry, and diffusion limited aggregation. Environ.Plann. 21,1447–1472.

Behan, R. W. 1990. Multiresource forest management: a parasigmatic challenge to professional forestry. Journal of Forestry. 88(4), 12-l8.

Bilsborrow, R. E., and H. W. O. Okoth Ogondo. 1992. Population-driven changes in land use in developing countries. *Ambio* 21:37–45.

Bingsheng K. 1996. Regional inequality in rural development. In:Garnaut R, Shutian G, Guonan M editors. The third revolution in the chinese countryside. Cambridge University Press, Cambridge. 245-258.

Breiman, L. 2001. Statistical Modeling. Statistical Science. Vol. 16, No. 3, 199–231.

Bürgi, M., Hersperger, Anna M. , and Schneeberger, Nina. 2004. Driving forces of landscape change - current and new directions. Landscape Ecology. Vol.19, N0o. 8, 857-868.

Carter, S.E., Jones, P.G.1993. Model of the distribution of cassava in Africa. Applied Geography. 13, 353-371.

Chapin, S. F., & Weiss, S. F. 1968. A probabilistic model for residential growth. Transportation
Research, 2, 375–390.

Chomitz, K.C. and Gray A.D. 1996. Roads, Land Use, and Deforestation: A Spatial Model Applied to Beliz. Oxford Journal, 10 485-512.

Clarke, K.C., Brass, J.A., Riggan, P.J., 1994. A cellular-automaton model of wildfire propagation and extinction. Photogrammetric Engineering and Remote Sensing 60(11), 1355–1367.

Clarke, K.C., Gaydos, L.J., 1998. Loose-coupling a cellular automaton model and GIS: long-term urban growth prediction for San Francisco and Washington/Baltimore. International Journal of Geographical Information Science 12(7), 699–714.

Clarke, K.C., Hoppen, S., Gaydos, L. 1997. A self-modifying cellular automaton model of historical urbanization in the San Francisco Bay area. Environment and Planning: Planning and Design. 24(2), 247 – 261.

Coleman, J.S. 1990. Foundation of social theory. The Belknap Press of Harvard University Press,

Collins, J., Kinzig, A., Grimm, N ., Fagan, W. ,Hope, D ., Wu, J., Borer, E.T. 2000. A New Urban Ecology. American Scientist, Vol. 88, 57-62.

Coughenour, M.B. 1991. Invited Synthesis Paper: Spatial Components of Plant-Herbivore Interactions in Pastoral, Ranching, and Native Ungulate Ecosystems. Journal of Range Management, Vol. 44, No. 6. 530-542.

Densham, P. J., 1991, Spatial decision support systems. In Proceedings of Geographical Information Systems: principles and applications, edited by D. J. Maguire, M. F. Goodchild and D. Rhind (London: Longman Scienti. c & Technical). 403–412.

Dramstad, W.E., Fjellstad, W. J. 2002. Development and implementation of the Norwegian monitoring programme for agricultural landscapes. Journal of Environmental Management. 64, 49–63.

Dramstad, W.E., Tveit, M.S., Fjellstad, W.J. 2006. Relationships between visual landscape preferences and map-based indicators of landscape structure. Landscape and Urban Planning. 78, 465–474.

Dunning, J. B., B. J. Danielson, and H. R. Pulliam. 1992. Ecological processes that affect populations in complex landscapes. Oikos. 65, 169-175.

Dunning, J. B., Jr., D. J. Stewart, B. J. Danielson, B. R. Noon, T. L. Root, R. H. Lamberson, and E. E. Stevens. 1995. Spatially explicit population models: current forms and future uses. Ecological Applications 5:3-ll.

Eetvelde, V.V., Antrop, M. 2004. Analyzing structural and functional changes of traditional landscapes—two examples from Southern France. Landscape and Urban Planning. 67, 79–95.

Fagan, W. F., Meir, E., Carroll, S. and Wu, J. 2004. The ecology of urban landscapes: modeling housing starts as a density-dependent colonization process**.** Landscape ecology. 16, 33-39.

Fischer, G., Sun L. 2001. Model based analysis of future land-use development in China Agriculture, Ecosystems and Environment. 85, 163-176.

Geist, H.J., Lambin, E.F. 2002. Proximate causes and underlying driving forces of tropical forestation. BioScience. 52, 143-150.

Geist, H.J., Lambin, E.F., 2002. Proximate causes and underlying driving forces of tropical deforestation. Bioscience 52(2), 143–150.

Gellerich, M., Baura, P., Kochb, B. and Zimmermann, N. E. 2007. Agriculture, Ecosystems & Environment. 118, 93-108.

Guisan, A., Zimmermann, N.E., 2000. Predictive habitat distribution models in ecology. Ecological Modelling 135(2/3), 147–186.

Heilig, G.K., 1994. Neglected dimensions of global land use change: reflections and data. Popul. Dev. Rev. 20 (4), 831–859.

Herold, M, Scepan, J., and Clarke, K. C. 2002. The use of remote sensing and landscale metrics to describe structures and changes in urban land uses", Environment and Planning A. 34, 1443-1458.

Hogeweg, P. 1988. Cellular Automata as a Paradigm for Ecological Modeling. Applied mathematics and computation 27:81-100.

Indovina, F., L. Fregolent and M. Savino. 2005. *L'esplosione della città.* Bologna.

Ingersoll, R. 2001. Verso Roadtown E-R. In *L.R.19/98, La riqualificazione delle aree urbane in Emilia-Romagna.* Bologna.

Ingersoll, R. 2002. Roadtown. *Firenze Architettura* 1&2.

Irwin, E.G. , Geoghegan, J. 2001. Theory, data, methods: developing spatially explicit economic models of land use change. Agriculture, Ecosystems and Environment. 85, 7–23.

Johnson, W.C. 1977. A mathematical model of forest succession and land use for the North Carolina Piedmont. Bull Torrey Bot. Sot. 104, 334-346.

Koning, G.H.J., Verburg, P.H., Veldkamp, A, Fresco, L.O. 1999. Multi-scale modelling of land use change dynamics in Ecuador. Agricultural Systems. 61, 77-93.

LaGro J.A., DeGloria S.D.1992. Land use dynamics within an urbanizing non-metropolitan county in New York State (USA). Landscape Ecology. 7, 275-289.

Lambin, E.F., 1994. Modelling deforestation processes: a review. TREES Series B. Research Report 1. European Commission, Brussels, EUR 15744 EN.

Lambin, E.F., Baulies, X., Bockstael, N., Fischer, G., Krug, T.,Leemans, R., Moran, E.F., Rindfuss, R.R., Sato, Y., Skole, D.,Turner, B.L. II, Vogel, C., 1999. Land-use and land-cover change
(LUCC):Implementation strategy. IGBP Report No. 48, IHDP Report No. 10, Stockholm, Bonn.

Lambin, E.F., Rounsevell, M.D.A., Geist, H.J. 2000. Are agricultural land-use models able to predict changes in land-use intensity?. Agriculture, Ecosystems and Environment. 82, 321-331.

Lambin, E.F., Turner, B.L. 2001. The causes of land-use and land-cover change moving beyond the myths. Global Environmental Change. 11, 261-269.

Li, H. and Reynolds J. F.1993. A new contagion index to quantify spatial patterns of landscapes. Landscape ecology. 8, 155-162.

Loucks, O.L., Ek, A.R., Johnson, W.C. and Monserud, R.A. 1981. Growth, aging and succession. In: Dynamic Properties of Forest Ecosystems. pp. 37-85. Edited by D.E. Reichle. Cambridge Univ. Press, Cambridge.

Mertens, B., Lambin, E. 1997. Spatial modelling of deforestation in southern Cameroon Spatial disaggregation of diverse deforestation process. Applied Geography.17, 143-162.

Munroe, D.K., Muller, D. 2007.Issues in spatially explicit statistical land-use/cover change (LUCC) models: Examples from western Honduras and the Central Highlands of Vietnam. Land Use Policy 24, 521–530.

Nelder, J.A.,Wedderburn, R.W.M. 1972. Generalized linear models. Journal of the Royal Statistical Society. Series A (General), Vol. 135, No. 3, 370-384

Neumann, J.V. 1966. Theory of Self-Reproducing Automata. Edited by A.W.Burks. University of Illinois Press Champaign, IL, USA.

Nichol, G.E., R.J. MacEwan, C. Pettit, 2J. Dorrough, H. Hossain, H. Suter, D. Cherry, C. Beverly, X. Cheng, V. Sposito, J. McNeill, A. Melland and A.Shanks. A Review Of Models Applicable To ‘Our Rural Landscape’. MODSIM conferences 2005 (Canberra).

Oliva, A.1938. Sistemazioni idraulico agrarie.

Risser, P.G., Karr, J.R. and Forman, R.T.T. 1984. Landscape Ecology: Directions and Approaches. Illinois Natural History. Survey Special Publ. 2, Champaign, Illinois.

Sereni,E. 1974. Storia del paesaggio agrario italiano. Laterza

Shugart H. H., West D.C. , 1980. Forest succession models, Bioscience, 30: 308-313.

Shugart, H. H., 1984, A theory of forest dynamics. The ecological implications of forest succession models. Springer.278pp.

Shugart, H.H., Jr. and Seagle, S.W. 1985. Modeling forest landscapes and the role of disturbance in ecosystems and communities. In: The Ecology of Natural Disturbance and Patch Dynamics. pp. 353-368. Edited by S.T.A. Pickett and P.S.White. Academic Press, New York. Ecology 1: 29-36.

Silva, E. A. and Clarke K. C. .2002. Calibration of the SLEUTH urban growth model for Lisbon and Porto, Portugal. Computers, Environment and Urban Systems ,Volume 26, Issue 6, November 2002, Pages 525-552.

Sklar, F.H., Costanza, R., 1991. The development of dynamic spatial models for landscape ecology: a review and prognosis. In: Turner, M.G., Gardner, R.H. (Eds.), Quantitative Methods in Landscape Ecology. Ecological Studies. 82, 239–288

Stéphenne, N., Lambin, E.F. 2001. A dynamic simulation model of land-use changes in Sudano-sahelian countries of Africa (SALU). Agriculture, Ecosystems & Environment. 85, 145-161.

Stoorvogel, J.J. 1995. he role of earthworms in the formation of sandy surface soils in a tropical forest in Ivory Coast. Geoderma 65, 135-148.

Tassinari, P., Benni, S., Torreggiani, D. 2007. Agriculture and Development Processes: Critical Aspects, Potential and Multilevel Analysis of Periurban Landscapes. Part II. Agricultural Engineering International: the CIGR journal. Manuscript MES 07 006.Vol. IX.

Tassinari, P., Carfagna, E, Benni, S., Torreggiani, D. 2008. Wide-area spatial analysis: A first methodological contribution for the study of changes in the rural built environment. Biosystem Engineering . Volume 100, number 3, 435-447.

Turner, M.G.1989. Landscape ecology: the effect of pattern on process. Annual Review of Ecology and Systematics. 20, 171-197.

Turner, M.G., Arthaud, Greg J., Engstrom, R. Todd , Hejl, Sallie J. 1995. Usefulness Of Spatially Explicit Population Models In Land Management. Ecological Application. 5(1), pp. 12-16 by he Ecological Society of America. T

Turner, M.G., Ross, R.H., Skole, D., 1993. Relating Land Use and Global Land Cover Change. IGBP report no.24, HDP report no. 5, p. 65.

Varni, A. 2007. Lo scorrere del paesaggio. Il trasformarsi della pianura romagnola dalla preistoria al ‘900. Edit Faenza.

Veldkamp, A., Fresco, L. O. 1997.Exploring Land Use Scenarios, An Alternative Approach based on actual land use" Agricultural Systems 55 1 – 17.

Veldkamp, A., Lambin, E.F. 2001. Predicting land use change. Agric. Ecosyst. Environm. 85, 1-6.

Verburg, P.H., Chenb, Y., Veldkamp, Tom (A.). 2000. Spatial explorations of land use change and grain production in China. Agriculture, Ecosystems and Environment. 82, 333-354.

Verburg, P.H., de Koning, G.H.J., Kok, K., Veldkamp, A., Bouma, J. 1999. A spatial explicit allocation procedure for modelling the pattern of land use change based upon actual land use. Ecological Modelling 116, 45–61.

Watson MK. 1978. The scale problem in human geography. Geografiska Annaler 60B: 36-47.

Weinstein, D.A. and Shugart, H.H. 1983. Ecological modelling of landscape dynamics. In: Disturbance and Ecosystems: Components of Response. pp. 29-45. Edited by H.A. Mooney and M. Godron. Springer-Verlag, New York.

Weisner, C., Cowen**,** D.J.1997.ACSM/ASPRS Annual Convention and Exposition Techical Papers:

White R, Engelen G, 1993, Cellular automata and fractal urban form: a cellular modelling approach to the evolution of urban land-use patterns. Environment and Planning A 25(8) 1175 – 1199

White, R., Engelen, G. 1997. Cellular automata as the basis of integrated dynamic regional modelling. Environment and Planning B: Planning and Design 24(2) 235 – 246.

Wolfram,S. 1984. University and complexity in cellular automata. Communications in Mathematical Physics. 96, 15-57.

Wu, F., Webster C.J. 1998. Simulation of land development through the integration of cellular automata and multicriteria evaluation. Environment and Planning. 25, 103 -126